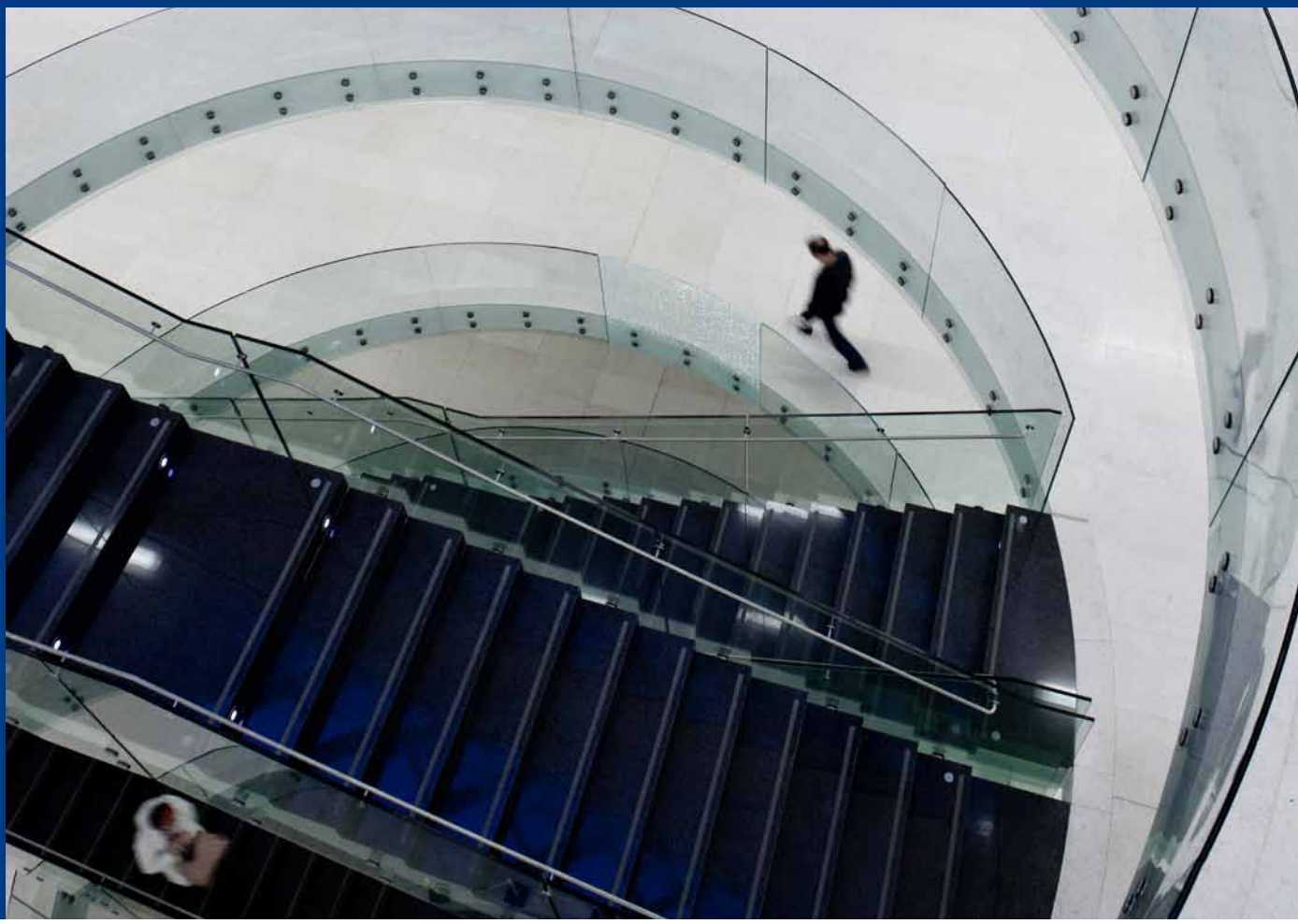
32nd International Conference

MME2014

Mathematical Methods in Economics

Conference Proceedings



September 10–12, 2014

Palacký University, Olomouc
Faculty of Science
Czech Republic

32nd International Conference

Mathematical Methods in Economics

MME 2014

Conference Proceedings

Olomouc, Czech Republic
September 10 – 12, 2014

Faculty of Science
Palacký University, Olomouc

Editors
Jana Talašová
Jan Stoklasa
Tomáš Talášek

Technical Editor
Tomáš Talášek

Cover Photo
Viktor Čáp

First edition
ISBN 978-80-244-4209-9

Contents

Pension reforms and welfare gains from eliminating business cycle in Poland

Jan Acedański[1]

Abstract. In the paper we study consequences of the changes in the pension system in Poland for the welfare gains from eliminating business cycle fluctuations. Using an overlapping generation model with aggregate fluctuations and exogenous idiosyncratic labour market risk we quantitatively show that raising the retirement age significantly reduces, whereas lowering the replacement rate from the public part of the pension system increases the business cycle costs. The impact of these changes is different as far as consumers with different skill level and wealth are concerned.

Keywords: welfare, business cycle, pension system, overlapping generations

JEL classification: E24,E21,E32,C63
AMS classification: 90C15

1 Introduction

The recent changes in the demographic structure in both developed and developing countries have brought serious consequences for their pension systems. Decreasing ratio of active population to retirees forces the policy makers to implement deep reforms of the labour markets and the pension systems. As a result in most European countries the retirement age has been increased whereas the replacement rates of public pensions are expected to drop significantly.

This is also the case in Poland. Before 2012 year the official retirement age in Poland was 65 years for men and 60 years for women. However in many cases due to privileges for some professions it was even lower so that in 2012 on average people exited the labour market at about the age of 60 [9]. Since 2012 year the retirement age has been gradually being increased to 67 years for both sexes. Similarly currently the gross replacement rate in Poland is estimated at about 50% [6], whereas simulations show that it will be impossible to keep it constant and it is expected to fall significantly to about 30% [8].

Since the pension provides a relatively stable, usually long-lasting income, it offers a good protection against the business cycle fluctuations for risk-averse consumers. As the retirement age increases and the public pensions drop down the protection becomes weaker. This is because both income from work and interest from individual capital stock that have to replace the lower pensions, are more volatile at the business cycle frequencies compared to the pensions. However on the other hand the discussed changes allow the government to significantly lower tax rates. Then consumers can accumulate more wealth to hedge against the business cycle fluctuations. As a result the effect of the reforms for the labour market cyclical variation is qualitatively uncertain.

In the paper we examine quantitatively to what extent the discussed consequences of the pension system reforms in Poland affect the costs of business cycle fluctuations for people of different age and skill level. To achieve this goal we build a general equilibrium overlapping generation model with aggregate fluctuations and idiosyncratic labour market risk. The parameters of the model are calibrated to match basic characteristics of the labour market and the pension system in Poland before and after the reforms as well as some business cycle stylized facts. To assess the potential welfare gains from eliminating the business cycle fluctuations we employ the approach proposed by Lucas [4]. We study two versions of the model: with and without the aggregate risk. In the former case the amount of the idiosyncratic labour market risk depends on the aggregate state of the economy, whereas in the latter case it is constant. Then we study a transition path from the economy without the aggregate risk to the economy with business

[1]University of Economics in Katowice, Department of Statistical and Mathematical Methods in Economics, ul. Bogucicka 14, 40-226 Katowice, Poland, jan.acedanski@ue.katowice.pl

cycle fluctuations. The welfare costs are defined as a percentage increase in individuals' consumption during the transition path needed to achieve the same lifetime utility from consumption as in the model without the aggregate risk. It should be stressed that the model does not take into account the changes in the demographic structure of the population.

The rest of the paper is organized as follow. In the next section we present the model. Then we discussed the model's calibration details. The fourth section contains the description of computational algorithms. Finally we show the results of the simulations and conclude the paper.

2 The model

We use a general equilibrium overlapping generation model with aggregate productivity shock and idiosyncratic risk following Storesletten, Telmer and Yaron [7]. However contrary to the cited work we do not focus on modelling exogenous individual income processes but rather assume that agents face exogenous labour market risk that is calibrated to match Polish data. Moreover we have a slightly different asset structure of the economy and allow for stochastic lifetime. We also assume exogenous labour supply as we exclude labour from the utility function. Below we describe the model with aggregate fluctuations. In the calibration section we highlight the differences between the model with and without the aggregate risk.

2.1 Description of the economy

The economy is populated by the continuum of finitely lived agents that differ in terms of age a, skill level s, employment status ϵ and wealth k. For simplicity we omit the time subscripts. The agents start at the age of 20, work for 40 years, then retire and live at most up to 100 years. Young agents either work or are unemployed. If employed, they supply $\xi_{s,a}l$ effective units of labour and get pre-tax income $\xi_{s,a}lW$, where W stands for the aggregate wage. Unemployed agents collect the unemployment benefits that are proportional to mean wage in the economy $\theta_u\overline{\xi}lW$, where θ_u is the unemployment replacement rate and $\overline{\xi}$ denotes mean labour efficiency across agents. Retirees receive pensions that are proportional to their last income from work $\theta_r\xi_{s,59}lW$, where θ_r represents the pension replacement rate. Moreover all agents collect interest from their wealth according the the aggregate interest rate R. We assume that agents who die leave their capital for new-born agents that replace them, but do not take into account leaving the bequest when facing their consumption-investment problems. This assumption helps to keep the model closer to the data without making the agents decision horizon infinite.

The production sector consists of one representative firm that hires capital and labour from agents producing a single consumption good according to standard Cobb-Douglas technology: $Y = ZK^{\alpha}L^{1-\alpha}$, where Z is stochastic aggregate productivity shock, $K = \int k_i di$ is aggregate capital stock and $L = l\int \xi_{s,a}\mathbf{I}(a_i \leq 59, \epsilon_i = e)di$ is aggregate effective labour supply. Aggregate wage and interest rate are given by: $W = (1-\alpha)ZK^{\alpha}L^{-\alpha}$ and $R = \alpha ZK^{\alpha-1}L^{1-\alpha}$.

There is also government in the model who imposes taxes on the income from work to finance the unemployment benefits and the pensions. The tax rate is set so the government budget is balanced every period.

2.2 Agents' decision problem

Every period an agent i of age a faces a following consumption-saving problem. Given her capital stock k, aggregate state of the economy Z_t and individual income d she has to decide how much of her current wealth to consume and how much to save to maximise her lifetime utility. The problem can be written in the form of Bellman equation:

$$V(k,a,\epsilon,s,K,Z) = \max_{c,k'}\{U(c) + \beta q_{a,a+1}\mathbb{E}\left[V(k',a+1,\epsilon',s,K',Z')|\epsilon,K,Z\right]\} \quad \text{s.t.} \tag{1}$$

$$k' = (1-\delta+R)k + d - c, \tag{2}$$

$$d = \mathbf{I}(a \leq 59, \epsilon = e)(1-\tau)\xi_{s,a}lW + \mathbf{I}(a \leq 59, \epsilon = u)\theta_u\overline{\xi lW} + \mathbf{I}(a > 59)\theta_r\xi_{s,59}lW, \quad k' \geq 0, \tag{3}$$

where V — value function, β — discount factor, $q_{a,a+1}$ — survival probability from age a to $a+1$, δ — capital depreciation, c — individual consumption, τ — tax rate, $U(c) = \frac{c^{1-\gamma}}{1-\gamma}$ — one-period CRRA utility function, γ — risk aversion parameter. For the notation transparency we also omitted individual subscripts and denoted next period with the prime. In (3) we impose borrowing constraints as the capital holding can not be negative.

The Euler equation describing an agent's decision rule regarding next period's capital when the borrowing constraint is not binding has the following form:

$$k' = (1 - \delta + R)k + d - \left[\beta q_{a,a+1} \mathbb{E}\left(\frac{1 - \delta + R'}{[(1 - \delta + R')k' + d' - k'']^{\gamma}}\right)\right]^{-\frac{1}{\gamma}}. \tag{4}$$

2.3 Stochastic structure of the economy

There are three stochastic shocks in the model. The aggregate productivity shock Z is represented by a two state Markov chain with transition matrix P_Z. The states of $Z = \{Z_b, Z_g\}$ represent recession and expansion periods in the economy. The individual employment shock ϵ is also modelled as a two state Markov chain with transition matrix $P_\epsilon(Z, Z', s, a)$ that depends on the current and future state of the economy as well as agent's skill level and age. Finally there is also a stochastic lifetime. For every generation a there is a fraction $1 - q_{a,a+1}$ of agents who die. It should also be noted that we do not allow agents to change their skill level.

3 Calibration

A period in the model corresponds to one year. Baseline calibration is presented in table 1. Share of capital in the production function α is set at 0.36 which is a standard choice for models of this class. The values of the capital depreciation rate $\delta = 0.032$ and discount rate $\beta = 0.98$ are calibrated jointly to match the interest rate of about 5% and average share of investment in output that is about 20% in Poland. In our baseline calibration we assume coefficient of relative risk aversion $\gamma = 5$. The compensation rates for agents who does not work equal $\theta_u = 0.15$ and $\theta_r = 0.5$, which roughly match the average values observed in Poland. As far as the skill level is concerned we have three group of agents in the model: low-, medium- and high-skilled which broadly correspond to people with primary, secondary and tertiary education level. Their average shares in Polish population equal 15%, 70% and 15% respectively.

Description	Parameter	Value
share of capital in the production function	α	0.36
capital depreciation rate	δ	0.032
discount rate	β	0.98
risk aversion	γ	5
unemployment benefit rate	θ_u	0.15
pension replacement rate	θ_r	0.5
share of low-, medium- and high-skilled workers	ω	{0.15, 0.7, 0.15}
individual labour supply	l	1

Table 1 Baseline calibration of the model

The individual efficiency factors ξ shown in table 2 are calibrated to match observed differences in wages according to the Structure of Earnings Survey conducted by Eurostat. The data show that earnings of low-skilled (primary-educated) workers are about 20% lower that for medium-skilled (secondary-educated), whereas for high-skilled (tertiary-educated) they are about 65% higher. Similarly it is documented that young people (20-29 years) earn about 20% lower wages than the rest. The differences for other age groups are small so we neglect them.

To calibrate Markov chain for the aggregate productivity we calculate log-deviations from HP-filtered trend of yearly data on Polish GDP covering period 1996-2013. This implies a symmetric business cycle where each phase lasts 3.75 years on average and values of the productivity shock $Z_b = 0.99$ during

	Skill level				Age	
	Low	Medium	High		20-29	30-59
ξ_s	0.8	1	1.65	ξ_a	0.8	1

Table 2 Calibration of the individual efficiency factors

recessions and $Z_g = 1.01$ for expansions. The transition probabilities P_ϵ are calibrated to match the average level $\bar{u}$ and duration u_l of unemployment in Poland during booms and recession for workers of different skill level and age based on the data from the Labour Force Survey. Since the data on mean level of unemployment are grouped into 5 year bins, the values for each age were interpolated. Due to scarcity of data on the unemployment length we assume that it is equal for all education levels and distinguish only two age groups: 20-39 and 40-59. We calculate the transition probabilities in the following way. The probability of staying unemployed: $p_{uu}(a, Z) = 1 - u_l^{-1}(a, Z)$. The probability of losing a job: $p_{eu}(a, s, Z) = [\bar{u}(a+1, s, Z') - \bar{u}(a, s, Z) \cdot (1 - p_{uu}(a, Z))] \,/\, [1 - \bar{u}(a, s, Z)]$, where $\bar{u}(a, s, Z)$ denotes mean unemployment rate for worker of age a, skill level s when the state of the economy is Z. Finally the survival rates $q(a, a+1)$ are taken from Polish unisex life tables from 2012 year.

For the model without the aggregate risk we set $Z = 1$. Moreover to calibrate the labour market transition probabilities we take mean level and duration of unemployment for all periods regardless of the business cycle phase. However we still take into account the differences across age and skill level.

4 Computational algorithm

Solving the model with the aggregate risk on the transition path we follow the general approximate aggregation algorithm proposed by Krusell and Smith [3] adapted to the overlapping generation models by Storesletten, Telmer and Yaron [7]. We assume that consumers are boundedly rational in the sense that they do not use the whole wealth distribution to predict next period aggregate wage and interest rate, which is an infinite dimensional object, but rather take into account only mean level of the aggregate capital K. They use simple autoregressive models to predict the next period stock of the aggregate capital:

$$K' = b_{0b} + b_{1b} K \text{ if } Z = Z_b, \qquad K' = b_{0g} + b_{1g} K \text{ if } Z = Z_g, \qquad b = [b_{0b}, b_{1b}, b_{0g}, b_{1g}]. \tag{5}$$

Krusell and Smith [3] proposed an iterative algorithm to find the coefficients b that are almost perfectly consistent with the actual dynamics of the aggregate capital in the model. It consists of the following steps:

1. Set initial values of the coefficients b^0.
2. For the given b^j find the individual's decision rules $c = c(k, a, \epsilon, s, K, Z, b^j)$, $k' = k'(k, a, \epsilon, s, K, Z, b^j)$ that solve the individual's consumption-saving problem (1)-(3).
3. Given the decision rules simulate the model for T periods starting from no aggregate risk stationary wealth distribution and compute a time path for the aggregate capital K.
4. Estimate the new autoregressive coefficients b_n^j using ordinary least squares.
5. If $|b^{j+1} - b^j| < \nu_{tol}$ then stop, otherwise update the coefficients b^j and return to step 2.

As far as the second step is considered we use a backward iteration method. Since agents do not derive utility from leaving bequests, the optimal decisions for the last generation are $k' = 0$ and $c = k + d$. Then we use this solution to find the decision of earlier generation. Here we employ the Euler equation (4) iteration approach proposed by Maliar, Maliar and Valli [5]. We discretize the state space for the individual and aggregate capital and look for the optimal next period's capital by iterating the Euler equation (4). More precisely with some initial decision rule k' and coefficients b^j we calculate the r.h.s. of (4) on the predefined grid and use it to update the decision rule. We iterate until convergence. To calculate the expectation term the transition probabilities are used as well as the decision rule of the next generation (to obtain k'' term).

In the third step we simulate the model. Here we do not perform full Monte Carlo simulations with some predefined number of agents. Instead we follow Heer and Maussner [2, p. 541–545] and simulate only aggregate productivity shock. Then for every period we analytically compute individual capital density functions for each skill level and age by taking advantage of the fact that with discretized individual capital level its dynamics is described by a Markov chain with transition probabilities depending on the current and the future state of the economy.

Finally it should be noted that for updating the parameters b as well as the individual decision rules k' we use the following updating rules: $b^{j+1} = \phi_b b^j + (1-\phi_b)b_n^j$ and $k'^{j+1} = \phi_k k'^j + (1-\phi_k)k_n'^j$.

In the model without the aggregate risk the aggregate capital is constant and so are the interest rate and the aggregate wage. As a result the model solution is easier. However rational agents still have to know the aggregate level of capital to accurately forecast interest rate and wage. Therefore we employ a similar solution strategy as in the model with the aggregate fluctuations assuming trivial capital forecasting rule $K = b$. And in such setup the decision rules obviously no longer depend on K.

For the individual capital holdings the decision rule are approximated on the interval [0, 50] with 100 grid points. Following Maliar, Maliar and Valli [5] we use a polynomially spaced grid of 7^{th} order that places more points near the borrowing constraint where the decision rule is far from linear. For the aggregate capital we use uniformly spaced 10-points grid which support is tuned for every parametrization we study. We simulate the law of motion for the aggregate capital for $T = 200$ periods. In this step we use equispaced interval for the individual capital stock with 200 grid points. The updating parameters equal: $\phi_b = 0.4$ and $\phi_k = 0.4$.

5 Results

The welfare gain from eliminating the business cycle fluctuations for an agent is defined as a percentage increase in individuals' consumption on the transition path needed to achieve the same lifetime utility from consumption as in the model without the aggregate risk. To compute the cost we simulate 100 consumption distribution paths in the two economies. Then for every simulated period we calculate the lifetime discounted utilities from the consumption for agents with median consumption who die at age 100 in that period. For every two lifetime utilities we solve for a constant consumption increase rate on the transition path, that makes the two lifetime utilities equal. Below we report mean values over 100 simulations. To assess the consequences of the pension reforms for the business cycle costs we consider economies in which the pension replacement rate drops to 25% as well as the case when the retirement age is increased from 60 to 67 years.

Model	All agents	Skill level			Wealth percentile	
	(median wealth)	Low	Medium	High	5	95
Baseline	2.03 (1.26)	1.82 (1.01)	2.19 (1.44)	2.38 (1.65)	-1.48 (-1.64)	3.14 (2.29)
$\theta_r = 0.25$	2.46 (1.69)	2.25 (1.48)	2.60 (1.83)	2.93 (2.11)	-0.38 (-0.72)	3.31 (2.43)
$RA = 67$	1.12 (0.59)	0.90 (0.37)	1.28 (0.77)	1.27 (0.79)	-1.33 (-1.45)	2.34 (1.65)
$\theta_r = 0.25$, $RA = 67$	1.51 (0.97)	1.37 (0.83)	1.61 (1.07)	1.65 (1.14)	-1.10 (-1.28)	1.14 (0.56)

Table 3 Maximum costs of business cycle (in percent of lifetime consumption) for median consumers (mean costs in parentheses)

The costs of business cycle under different pension system characteristics are presented in table 3. In every case we report two values: the maximum lifetime costs which is usually associated with an agent who is born at the beginning of the transition path and dies in 80^{th} period, as well as the mean costs for agents who die at age 100 between 60 and 120 year of the transition period. The first column contains costs for all agents with median wealth level. Then we present also results for agents with different skill level as well as wealth.

In baseline version of the model maximum cost of the business cycle fluctuations is about 2% of lifetime consumption and on average it is 1.3%. It also increases with the consumers skill level. It is also worth mentioning that for the poorest consumers the cost is negative, which means that they are better

in the economy with fluctuations. This is because relatively large part of their income comes from the social security system that, due to higher capital accumulation, is more generous in the economy with the business cycle fluctuations.

With lower pension replacement rate the costs rises significantly to 2.5% and 1.7% respectively. In relative terms the reform would boost the costs by 25%. Again the increase is higher for high-skilled workers. Interestingly it affects especially the poorest people and is virtually harmless for the wealthiest consumers. Increasing the retirement age on the other hand halves the costs to 1.1% and 0.6% respectively regardless of the skill level. However contrary to the previous case the rise of the retirement age is beneficial mainly for the wealthiest consumers but do not matter for the poorest. Finally considering simultaneously the lower replacement rate and higher retirement age the maximum cost decreases to about 1.5% and mean cost to 1% of the lifetime consumption. The drop is highest for high-skilled and wealthiest consumers.

6 Conclusion

In the paper we quantitatively studied the consequences of the pension system reforms for the cost of the business cycle fluctuations in Poland. The results of our simulation study suggest that raising the retirement age significantly reduces, whereas lowering the replacement rate from the public part of the pension system increases the business cycle costs. Overall the two changes reduce the maximum cost from 2% to 1.5% of lifetime consumption for the median consumer. That means that the benefits from increasing individual capital accumulation possibilities under the new pension scheme outweigh the negative effects of less generous pensions. This is the case especially for high skilled and wealthiest consumers, although they still suffer the most from the business cycle fluctuations on labour market.

Acknowledgements

The author thanks two anonymous referees for helpful comments and discussion. The author is solely responsible for any remaining errors. The support from the Polish National Science Center grant DEC-2011/01/D/HS4/03430 is kindly acknowledged.

References

[1] Epstein, N., and Macchiarelli, C.: *Estimating Poland's potential output: A production function approach*. IMF Working Paper WP/10/15, 2010.

[2] Heer, B., and Maussner, A.: *Dynamic General Equilibrium Modeling. Computational Methods and Applications*, Springer, Berlin, Heidelberg, 2009.

[3] Krusell, P., and Smith, A. A.: Income and wealth heterogeneity in the macroeconomy, *Journal of Political Economy* **106** (1998), 867–896.

[4] Lucas, R. E.: *Models of Business Cycles*. Basil Blackwell, Oxford 1987.

[5] Maliar, L., Maliar, S., and Valli, F.: Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm, *Journal of Economic Dynamics and Control* **34** (2010), 42–49.

[6] *Pensions at a Glance 2013. OECD and G20 Indicators*. OECD 2013.

[7] Storesletten, K., Telmer, C., and Yaron, A.: Asset pricing with idiosyncratic risk and overlapping generations, *Review of Economic Dynamics* **10** (2007), 519–548.

[8] *The 2012 Ageing Report. Economic and Budgetary Projections for the 27 EU Member States (2010-2060)*. European Commission, 2012.

[9] *Transition from Work into Retirement in 2012*. Central Statistical Office, Warsaw, 2013.

Effects of Interventions and Outliers on Mean Function and Volatility in Currency Exchange Rates

Václav Adamec[1]

Abstract. Stochastic time series frequently incorporate information from the outside world. The exterior information may sometimes generate extreme observations at one or several positions, which behave as outliers. Presence of outliers negatively influences estimation of parameters in time series models. Especially, the error variance may be poorly estimated. In other situations, time series may be affected by external shocks, i.e. interventions, as a result of natural disasters, introduction of new technologies, altered legislation or due to realization of political decisions. The interventions may alter the mean function and (or) volatility in time series. This study demonstrates use of ARIMAX models in detection and testing of additive and innovative outliers together with estimating the dynamic effects and effects of known external interventions in currency exchange rates. Also, GARCH models of conditional volatility were constructed to show evidence of a significant reduction in volatility in the post intervention period in CZK/EUR and CZK/GBP rates. It is hypothesized that knowledge of the intervention in the foreign exchange market increased awareness about future value of exchange rates and led to lower volatility.

Keywords: ARIMAX model, intervention analysis, outlier detection, exchange rate, volatility model.

JEL Classification: C22, C53, C58
AMS Classification: 62M10

1 Introduction

It is widely acknowledged, that Board of the Czech National Bank (CNB) decided on its regular meeting held on November 7, 2013, that effective immediately the central bank would launch intervention on the foreign exchange (FOREX) market against Czech koruna with the objective to weaken the CZK/EUR exchange rate (ER) by approximately one koruna per Euro. This move by the CNB was justified by the fact, that central bank could not further use its conventional tool, i.e. decrease interest rates, which were at "technical zero" level of 0.05% in Discount rate and 0.25% in Lombard rate. Also, at that time, there were considerable risks of deflation and the Czech economy showed only weak signs of recovery from the retreating recession. In this situation, cheaper koruna, could boost export of industrial products, which constitutes a significant component of Czech GDP growth, while mildly increasing price of imports. Emission of assets denominated in CZK increased domestic monetary base and was followed by raise in the general price level, since the Czech economy is small and open and it depends strongly on foreign markets. Volume of the intervention was initially estimated at CZK 80 bil., that were spent on the day of CNB involvement to acquire assets denominated in Euro.

The Czech central bank broadly communicated reasons and objectives of the intervention on the foreign exchange market with the general public. The bank further announced that it was prepared to maintain weaker CZK/EUR rate for at least 15 months or until deflation threats disappear. Full rationale behind the central bank decision to intervene against Czech currency is available at www.cnb.cz.

Objective of this study is to quantify impact of November 7, 2013 intervention against Czech koruna on the foreign exchange market by the CNB with focus on the adjustment of the mean levels of daily CZK/EUR and CZK/GBP exchange rates. We expect that despite the intervention be aimed directly at CZK/EUR rate, there can be a correlated response in exchange rates with other currencies. Further, we intend to describe impact of unusual data points, i.e. additive and innovative outliers and other hypothesized systematic effects on the quality of fit of the ARIMAX model, which combines autoregressive process and (or) process of moving averages described by Box and Jenkins [2] with exogenous fixed regressors. Information about the CNB intervention was communicated in detail with the public. For this reason, we also propose to investigate consequences of this operation on volatility in CZK/EUR and CZK/GBP exchange rates.

[1] Mendel University, School of Business and Economics, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: vadamec@mendelu.cz.

2 ARIMAX models with external covariates

Provided non-stationary time series can be modeled by stationary ARMA(*p*, *q*) after one or two rounds of differencing, the process is integrated ARIMA(*p*, *d*, *q*), where *d* denotes the order of integration. It is assumed, that differencing of non-stationary process with nonzero mean produces a stationary process with mean zero. ARIMA(*p*, *d*, *q*) can be written

$$\phi_p(B)(1-B)^d y_t = \theta_q(B)e_t\,, \tag{1}$$

where $\phi_p(B) = (1-\phi_1 B - ... - \phi_p B^p)$, $\theta_q(B) = (1-\theta_1 B - ... - \theta_q B^q)$, $\nabla^d y_t = (1-B)^d y_t$ denotes regular differencing of y_t, applied d times and e_t is random disturbance. B indicates backshift operator. Arlt and Arltová [1] provide general guidelines to determine order of model polynomials from typical patterns of sample ACF and PACF. Estimation of ARIMA(*p*, *d*, *q*) is possible by Maximum Likelihood (ML) or Conditional Least Squares (CLS) [11]. The random component is expected to comply with the assumptions of white noise and normality.

General principles of ARIMA(*p*, *d*, *q*) models were originally formulated by Box and Jenkins [2]. Incorporating external information from the outside world can further extend ARIMA models. The exterior information may in some cases account for unusual observations at one or several time positions. The unusual data points frequently behave as outliers. Presence of outliers negatively influences estimation of model parameters. Especially, the error variance is affected by overestimation [7].

In other situations, time series may be influenced by external shocks, i.e. interventions, as a result of natural disasters, introduction of new technologies, altered legislation or due to implementation of political decisions. Nature of the outside shocks upon the target series is strictly exogenous; their impact is either temporary or they may change mean function for a long time. ARIMAX models are therefore introduced to account for external covariates and at the same time explain the correlation structure of time series. Box and Tiao [3] published a pioneering paper that laid foundations of time series models with regression covariates.

With regards to unusual observations, two types can be recognized. Additive outlier (AO) occurs at time *T*, if the underlying process is locally lifted by some amount, i.e. $Y_t^* = Y_t + m_t$. Y_t denotes the unaltered process, $m_t = \omega_A P_t^T$ is the change of the mean function, ω_A is additive constant and P_t^T a pulse indicator variable, taking value of one at time T and zero otherwise. AO is detected by scanning the data and computing test statistic $\lambda_{A,T} = \tilde{\omega}_{T,A}/(\rho\sigma) \sim N(0,1)$ for every time position. $\tilde{\omega}_{T,A}$ is estimated shift, $\rho = (1+\pi_1^2+\pi_2^2+...+\pi_{n-T}^2)^{-1/2}$ and σ error standard deviation. Bonferroni correction for multiple tests provides, that overall probability of type I. error is at most 5% [7].

Innovative outlier (IO) occurs at time *T*, if the random innovation e_t, is shifted by some constant. In consequence, $e_t^* = e_t + \omega_I P_t^T$, where ω_I denotes the shift. The affected process is then $Y_t^* = Y_t + \omega_I \psi_{t-T}$, where $\psi_0 = 1$ and $\psi_j = 0$, $\forall j < 0$. IO affects all observations at time T and further, but with decaying effect. The test statistic for IO is calculated $\lambda_{I,T} = a_t/\sigma \sim N(0,1)$, where a_t denotes the potentially disturbed residual. IO is detected, if $|\lambda_{I,T}| > z_{1-\alpha/(2n)}$ to protect against exceeding the overall 5% rate [7].

External shocks (interventions) at time T can be modeled in a similar way. We assume a general model $Y_t = V_t + m_t$, where m_t is change of the mean function and V_t some innate process described by ARIMA. Depending on type of the intervention, the model includes a step indicator variable S_t^T, which takes value of one, if $t \geq T$ and zero otherwise, provided change of the mean function is long-term. On the other hand, if the change is short-lived, a pulse indicator variable P_t^T is introduced. Box and Tiao [3] present multiple variants of m_t, depending on the type of the exterior shock. The frequented forms of mean function change, assuming use of pulse function and lagged immediate change, include

- $m_t = \omega B P_t^T$, an immediate shift by ω,
- $m_t = \frac{\omega B}{1-\delta B} P_t^T$, an immediate shift by ω followed by first-order autoregression shift with decay parameter $\delta < 1$ until steady state change $\omega/(1-\delta)$ is attained, and

- $m_t = \frac{\omega B}{1-B} P_t^T$, a linear change with slope ω. Note, that $\delta = 1$.

3 Material and methods

To investigate effects of exterior regressors upon a stochastic time series, we used daily time series of CZK/EUR and CZK/GBP exchange rates from January 2013 until April 2014 (T = 336). The data segment before the intervention had 216 observations; on the intervention day and after, there were 120 observations available. Descriptive statistics and statistical tests for shape of the distribution (p-value) and first-order autocorrelation in segments before and after intervention are shown in Tab. 1. Evidently, the intervention led to increase in mean rate, significant drop in variability and in CZK/EUR rate it changed shape of the distribution to become more left-skewed and peaked.

segment	mean	sd	minimum	maximum	skewness	kurtosis	ρ_1
Jan 2 to Nov 6	25.743	0.174	25.225	26.120	-0.557 (0.033)	0.354 (0.234)	0.899
Nov 7 to Apr 30	27.418	0.132	26.850	27.720	-1.609 (0.000)	4.350 (0.000)	0.863
Jan 2 to Nov 6	30.242	0.425	29.052	31.450	-0.146 (0.554)	0.530 (0.115)	0.917
Nov 7 to Apr 30	33.037	0.324	31.973	33.623	-0.819 (0.023)	0.428 (0.243)	0.902

Table 1 Descriptive statistics for daily CZK/EUR (upper) and GBP/CZK (lower) exchange rates before and after the CNB intervention.

Following visual inspection of data plots, the ER series were subjected to ADF test of unit root [13] with level constant in the auxiliary regression. ($\varphi - 1$ = -0.005, p = 0.812; $\varphi - 1$ = -0.038, p = 0.121). Output of the non-stationarity test was supported by correlogram. Model of integrated ARIMA was then constructed and estimated by ML. Detection of additive and innovative outliers was completed by the method of Chang et al. [6] with application of Bonferroni rule and robust estimation of sigma. For a detected outlier, type of outlier, location of disturbed observation and empirical z-statistic were produced.

Integrated model of ARIMAX was constructed in the next step to incorporate information about position of statistically proven additive and innovative outliers and known date of CNB intervention. We assumed that the instantaneous effect of the intervention ω was positive and it was succeeded by first-order autoregressive change with estimated decay parameter δ. Further, we tested exogenous effects of special holidays on the systematic component, i.e. days following instantly after official state or religious holidays. According to this definition the holiday days nonetheless excluded Mondays. It was also hypothesized, that categorical variable of weekdays might contribute to the mean level of exchange rates. For this reason, four dummy variables were prepared with Fridays set as the reference level.

Estimated variant specifications of the ARIMAX model were compared by AIC criterion and parameter significance. Likelihood-ratio (LR) test was considered to assess simultaneous significance of parameter subsets, when required. Estimated residual component was checked for white noise properties during model diagnostics. Standard error of a nonlinear parametric function was obtained by Taylor series approximation, i.e. the Delta method [4]. Statistical analyses, especially models of the mean function were prepared with *R*-software [10] with libraries *TSA* [5] and *xts* [12]. Gretl 1.9.14. software, mainly the *gig* package [9] was used for estimating models of conditional volatility GARCH [1].

4 Results and discussion

Measures of the central bank aimed at influencing the foreign currency market are considered questionable from numerous viewpoints. Generally, they are rather nonstandard actions by the central bank focused not only on changing the mean level of the exchange rate, but also on moderating the volatility. In the short run, changing the mean level can be achieved, but the results depend mostly on circumstance of the intervention. Provided the central bank communicates its operations openly with the markets and the markets are totally efficient, the intervention can be useful in terms of maintaining the targeted level of ER and in addition, it may have calming influence on volatility [8]. Covert bank operations, on the other hand, bring more uncertainty to the markets and may result in exactly the opposite reaction, i.e. increase volatility. Because of potentially dubious impact of central bank interventions, we present the results separately for the mean function and volatility.

4.1 Models of the mean function

Based on exploration of the data and preliminary tests, a regression model of non-seasonally differenced CZK/EUR was constructed. It included binary indicator variables corresponding to weekdays, outliers detected by the method of Chang et al. [6] and two parameters describing the effect of CNB intervention on November 7, 2013: instant change represented by ω and parameter of subsequent autoregressive change $\delta < 1$. Because differenced data were modeled, a binary pulse variable P_T was applied instead of step variable S_T. First round of model diagnostics failed to establish evidence, that holidays had systematic effect on the mean level of both rates. For this reason, this regressor was omitted in the final model. The model in symbolic form is shown in (2).

$$(1-B)y_t = \sum_{j=1}^{4} \alpha_j D_{jt} + \beta_1 D_{IO-27} + \beta_2 D_{AO-110} + \beta_3 D_{IO-225} + \frac{\omega}{1-\delta B} P_{217} + e_t \,. \tag{2}$$

In consequence of CNB intervention, there was an imminent increase of the mean exchange rate by 1.07 koruna per Euro. Czech koruna thus weakened against Euro by 4.1 % with most of the change realized on the intervention day. In the following days, this change was further elevated to final steady state level 1.2 ± 0.084, i.e. 4.7 % increase relative to the pre-intervention boundary. The second gradual increase was significant (p = 0.046) and was realized in a short time.

The integrated regression model accounted for three outliers. Two of them, indicated limited, but significant strengthening of koruna, most likely as a result of some positive news made available to the public. It is known, that every 7-th day of the month the Czech Statistical Office publishes performance indicators related to the recent month for the industry, construction and retail sectors and foreign trade. Negative sign of the estimates can be easily explained by continuing positive balance of the foreign trade account. The third innovative outlier was dated 12 days after the official intervention was launched. It meant limited weakening of the koruna due to activities of the traders on the market, not excluding the CNB. Central banks usually do not publish data about their contemporaneous presence on the market [8]

Weekdays had statistically significant effect on the mean rate, although from the practical standpoint, magnitude of the differences was low. The largest hikes were on Mondays or Wednesdays, which is in accordance with our expectation. Standard deviation of the random noise was estimated $\hat{\sigma} = 0.0527$. It is significantly lower than $\hat{\sigma} = 0.0849$ received from integrated model excluding outer regressors. Estimated coefficients of model of differenced CZK/EUR exchange rate are shown in Tab. 2.

parameter	estimate	std. err.	z-statistic	p-value	weekday / date
α_1	0.0138	0.0065	2.1304	0.0331	Monday
α_2	0.0041	0.0064	0.6345	0.5258	Tuesday
α_3	0.0134	0.0065	2.0477	0.0406	Wednesday
α_4	-0.0050	0.0065	-0.7710	0.4407	Thursday
β_1	-0.4600	0.0531	-8.6651	0.0000	February 7, 2013
β_2	-0.2050	0.0527	-3.8908	0.0001	June 7, 2013
β_3	0.2209	0.0531	4.1620	0.0000	November 19, 2013
ω	1.0699	0.0531	20.154	0.0000	November 7, 2013
δ	0.1087	0.0489	2.2260	0.0260	November 7, 2013

Table 2 Estimated parameters of regression model of differenced CZK/EUR.

In CZK/GBP data, an ARIMAX model was fitted. It included external regressors weekdays, a single outlier dated February 7, 2013, two parameters describing the CNB intervention on November 7, 2013 and parameter of first-order moving averages. Holiday days did not test to be statistically significant. The model in symbolic form is in equation (3); parameter estimates are provided in Tab. 3.

$$(1-B)y_t = \sum_{j=1}^{4} \alpha_j D_{jt} + \beta_1 D_{IO-27} + \frac{\omega}{1-\delta B} P_{217} + (1-\theta_1 B)e_t \,. \tag{3}$$

Although the CNB intervened directly against CZK/EUR rate, there was also an indirect correlated impact on CZK/GDP rate on the day of the intervention. Koruna weakened against British pound instantly by 1.56 (5 %). Its impact was nonetheless felt for a prolonged period after the intervention due to parameter δ close to unity. We observed weekdays with significant peaks on Thursdays. The model confirmed, that the news, which

affected CZK/EUR on February 7, 2013 also strengthened CZK/GDP rate by 0.54 koruna. Error standard deviation was $\hat{\sigma} = 0.1413$. It is significantly lower than $\hat{\sigma} = 0.1712$ from ARIMA without external regressors.

parameter	estimate	std. err.	*z*-statistic	*p*-value	weekday / date
α_1	0.0081	0.0159	0.5077	0.6117	Monday
α_2	-0.0084	0.0183	-0.4589	0.6463	Tuesday
α_3	0.0254	0.0187	1.3574	0.1747	Wednesday
α_4	0.0374	0.0159	2.3515	0.0187	Thursday
β_1	-0.5379	0.1062	-5.0642	0.0000	February 7, 2013
θ_1	-0.1094	0.0536	-2.0410	0.0413	×
ω	1.5626	0.1413	11.057	0.0000	November 7, 2013
δ	0.9976	0.0038	263.12	0.0000	November 7, 2013

Table 3 Estimated parameters of ARIMAX model of CZK/GBP.

4.2 Models of volatility

Exploration of variances from data before and after CNB intervention suggests, that volatility in ER decreased shortly after the intervention. To verify this empirical observation, GARCH(0,1) model was constructed with external covariates, which included weekdays (D), days markets reopened after special holidays (H) and step intervention dummy (I) separating both segments. Model of the conditional variance is given in (4).

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \sum_{j=1}^{4} \gamma_j D_{jt} + \beta_1 I + \beta_2 H \tag{4}$$

There is significant evidence in the estimated model of conditional heteroskedasticity that volatility in random disturbances decreased in both time series in the period after the CNB intervention (see Tab. 4). Knowledge about the central bank involvement and its objectives provide credible information about likely future value of the ER. This awareness, together with the reality that the CNB has rather unlimited resources at its disposal to fulfill its declared commitment to increasing CZK/EUR rate, act as a likely stabilizer of ER volatility and reduce prospects of unpredictable fluctuations or speculative attacks. The stabilization effect on volatility is clear in CZK/EUR rate, although in CZK/GBP it was smaller, but significant (see Fig. 1). This finding is in agreement with assumption made by Dominguez [8] about the way secret and reported interventions influence volatility in the foreign currency market.

parameter	estimate × 100	std. err. × 100	*z*-statistic	*p*-value	estimate × 100	std. err. × 100	*z*-statistic	*p*-value
α_0	0.3661	0.0546	6.7100	0.0000	1.7139	0.3267	5.2460	0.0000
α_1	1.1458	4.4717	0.2562	0.7978	11.147	6.4490	1.7280	0.0839
γ_1	-0.0364	0.0556	-0.6541	0.5130	0.2492	0.4631	0.5380	0.5906
γ_2	-0.0824	0.0480	-1.7180	0.0858	-0.0318	0.3696	-0.0861	0.9314
γ_3	-0.0927	0.0448	-2.0700	0.0385	0.8744	0.4791	1.8250	0.0680
γ_4	0.0473	0.0758	0.6243	0.5324	0.2456	0.4005	0.6131	0.5398
β_1	-0.2030	0.0321	-6.3340	0.0000	-0.6544	0.2908	-2.2500	0.0244
β_2	0.6343	0.2975	2.1320	0.0330	×	×	×	×

Table 4 Estimated coefficients of volatility models of CZK/EUR (left) and CZK/GBP (right).

In the volatility model of CZK/EUR, we have noticed short-term increase in variability in days, when markets reopened after a special holiday. In CZK/GBP rate, on the other hand, we did not see any significant shifts in volatility in these days. Among weekdays, Wednesdays had the most moderating influence on volatility.

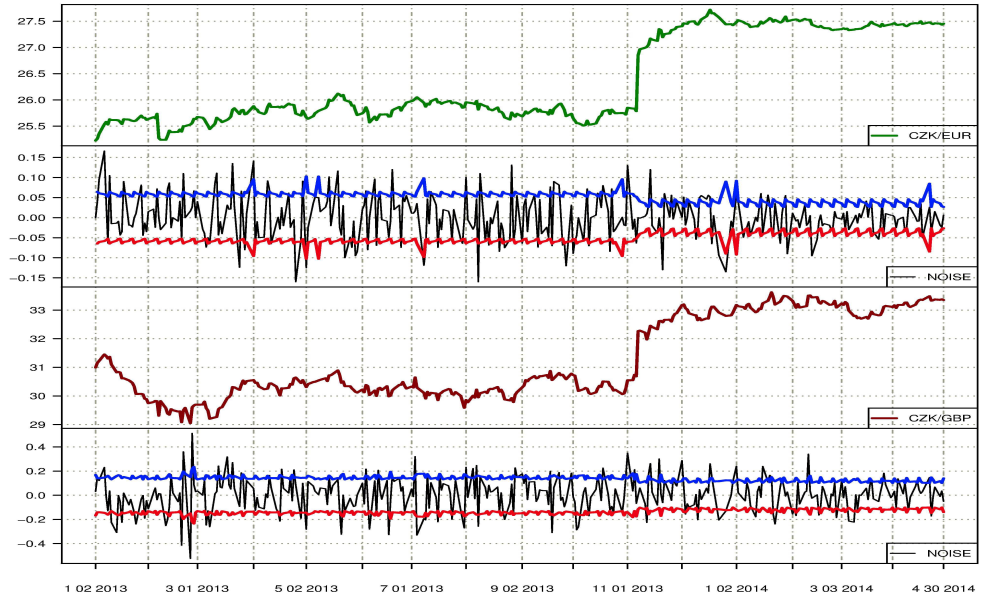


Figure 1 Observed data, estimated residuals and conditional standard deviation.

5 Conclusions

This paper brings results of ARIMAX model application to daily CZK/EUR and CZK/GBP exchange rates to account for external sources of variation or unusual data points. Evidence was found, that the CNB intervention on November 7 and weekdays influenced the mean level. Better quality of fit in ARIMAX model cuts positive bias in estimates of variance; improves model specification, and leads to more reliable statistical tests and predictions. GARCH models indicate, that the CNB intervention caused a significant fall in volatility in the post intervention period, especially in CZK/EUR rate. This finding may be explained by the fact, that markets were informed by the CNB of the intervention shortly after it begun. This clear signal increased awareness about future value of ER and contributed to larger stability. It also provides indirect evidence, that current FOREX market is capable of quickly accommodating available information in formation of prices and it is therefore efficient.

References

[1] Arlt, J., and Arltová, M.: *Ekonomické časové řady – vlastnosti, metody modelování, příklady a aplikace.* Grada Publishing a.s., Praha, 2007.
[2] Box, G. E. P., Jenkins, G. M., and Reinsel, G. C.: *Time series analysis, forecasting and control.* Prentice Hall, New York, 1994.
[3] Box, G. E. P., and Tiao, G. C.: Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, **70** (1975), 70–79.
[4] Casella, G., and Berger, R. L.: *Statistical inference.* Duxbury Press, 2002.
[5] Chan, K. S., and Ripley, B.: TSA: time series analysis. R package version 1.01. 2012, http://cran.r-project.org/ package=TSA.
[6] Chang, I. H., Tiao, G. C., and Chen, C: Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*, **30** (1988), 193–204.
[7] Cryer, J. D., and Chan, K. S.: *Time series analysis with applications in R.* Springer, New York, 2008.
[8] Dominguez, K. M.: Central bank intervention and exchange rate volatility. *Journal of International Money and Finance*, **17** (1998), 161–190.
[9] Lucchetti, J., and Balietti, S.: The gig package version 2.13. http://ricardo.ecn.wfu.edu/gretl/cgi-bin/.
[10] R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014, http://www.R-project.org/.
[11] Ruppert, D.: *Statistics and data analysis for financial engineering.* Springer, New York, 2011.
[12] Ryan, J. A., and Ulrich, J. M.: xts: extensible time series. R package version 0.9-7. 2014, http://cran.r-project.org/package=xts.
[13] Said, S. E., and Dickey, D. A.: Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, **71** (1984), 599–607.

The measure of separation of components in final mixtures of probability densities

Jitka Bartošová[1], Vladislav Bína[2]

Abstract. The attempts to model the large random samples frequently end with the finding that there does not exist a single appropriate distribution which will be both known and analytically feasible even in case of application of the transformation from some flexible class (e.g, the Box-Cox class). Therefore, it is reasonable to employ as a tool for modeling of the empirical density the finite mixtures of probability densities. The mixture modeling leads to the distribution with many different properties which can be easily estimated using the EM-algorithm.
The particular components of finite mixtures of income density functions calculated, e.g., in the R environment using the mclust procedure (EM/BIC algorithm) can be reckoned as latent clusters. But simultaneously the question arises, to what extent such components can be considered as clusters and how "objective" the assignment of the objects (households or individuals) to the clusters actually is. We will approach the problem from the perspective of the separation measure of the discovered clusters. The paper extends the Hennig's approach and concerns the question: to what extent the resulting components can be considered as clusters. The methodology is applied to the data file from the sample survey of incomes and living conditions of households EU-SILC 2010

Keywords: finite mixture of densities, mclust, separation, cluster.

JEL Classification: C44
AMS Classification: 62E17

1 Object clustering

One of the approaches leading to the simplification of complex structures with many categories is grouping (clustering) of objects to the isolated (separated) subgroups with similar properties. The ultimate aim of cluster analysis is to classify objects into the mutually disjoint subsets (clusters).

"*Basic methods of cluster analysis (CA) enable the identification of groups (clusters) of similar categories of one variable based on the categories of other variable or provide a possibility to find the relations between categories of both variables.*" (Řezanková, 2007)

1.1 Clustering methods

In general, two types of clustering methods are distinguished:

- hierarchical (agglomerative – starting with single elements and joining them into the clusters and divisive – starting with whole data set and dividing it into the clusters);
- non-hierarchical (optimization).

Hierarchical clustering does not demand any preliminary knowledge about the number of clusters and provides a possibility to create and display data structure (using dendrogram). On the other hand, non-hierarchical clustering algorithms need to set the number of clusters in advance. The success of the decomposition depends on the quality of an initial division, which sorts the units according to some complementary information or randomly. In the subsequent steps the objects are relocated according to their distance from the cluster centers. A usual advice says that the initial decomposition can be obtained by the hierarchical clustering and then further refined by (iterative) use of non-hierarchical methods.

Clustering and classification tasks are based on the measuring of distance between objects (distance-based clustering) or on the probability of the membership of the object in a latent cluster – component of a mixture model (model-based clustering). Successively, a bunch of methods were developed, which differ in the approach to the clustering. The above mentioned classification corresponds to two basic categories of clustering methods:

[1,2] University of Economics/Faculty of Management, Department of Information Management, Jarošovská 1117/II, 37701 Jindřichův Hradec, {bartosov,bina}@fm.vse.cz..

- clustering methods based on the distance measures, assigning particular objects unambiguously into the classes (for details see Řezanková, Húsek and Snášel, 2009);
- probabilistic (Bayesian) methods based on the decomposition of the mixture, assigning to the object the probabilities of membership to particular latent classes (components of mixture model).

1.2 Estimation of the number of clusters

The Bayesian clustering methods require to set the number of clusters (components K of the mixture) in advance. But it is unknown a priori. It has to be estimated expertly or using some of the well-known (penalized) information criteria. Among the most frequently used rank

- Akaike information criterion

$$AIC = 2m - 2\ln(\textit{maximized } L) \tag{1}$$

  where m is the number of unknown parameters and *maximized* L is maximum of the likelihood function for given model;
- Bayesian information criterion

$$BIC = m\ln(n) - 2\ln(\textit{maximized } L) \tag{2}$$

  where n is the number of all observations (see Akaike, 1977, Schwarz, 1978).

The *AIC* criterion stems from the concept of information entropy and its value indicates the relative loss of information in the model simplification of reality. The *BIC* criterion employs the covariance matrix for the search of optimal model (for more details see Fraley and Raftery, 2003). Criteria do not contain information concerning the accuracy of the model, i.e., they do not serve for the testing of model's statistical significance. But the criteria allow us to compare pairs of models and evaluate the difference in information loss when using the first model in comparison to the second.

With the growing number of parameters, in general, the accuracy of model grows. Therefore, theoretically, the best model would be the one with infinite number of components. But indeed, it will be the most complex one. The attempt to find an optimal model as a tradeoff between the maximum accuracy and simplicity of the model leads to the penalization in the information criteria. Thus it is a component that brings into the decision making the "penalty" for the higher number of parameters, i.e., the growth of its complexity. Akaike criterion penalizes the number of parameters less than the Bayesian criterion.

1.3 Optimal estimation of the mixture model in R software

Bayesian clustering is performed in the well-known *EM*-algorithm (Expectation-Maximization) which is based on the assumption of mixture character of data. The clustering is based on the estimation of mixture component parameters using the iterative procedure for the search of the local maximum of the respective likelihood function. The optimization process finds a finite mixture (of probability densities or regression models for particular components] able to model given empirical situation with the maximal likelihood. Thus the primary aim of the *EM*-algorithm is not the sorting of objects into the latent classes (clusters) but the construction of an optimal model of the empirical distribution or dependence of observed variable on different quantitative or qualitative factors.

For the detection and estimation of mixture model components often the *mclust* package (viz Fraley and Raftery, 2006) is used implemented in the R statistical software. Algorithm contained in the *mclust* package performs an *EM*-algorithm for multidimensional normal density mixtures. The *Mclust* function searches for na optimal mixture model based on different covariance structures and various numbers of clusters. For the optimal estimation of the cluster count, the Schwarz's Bayesian information criterion (*BIC*) is used. The triplet of characters (E, V, I) refers to the different assumption concerning the distribution, scope, shape and orientation of clusters. The *mclust* procedure provides the following variants of clustering:

- EII – spherical, equal volume, equal shape, not oriented;
- VII – spherical, unequal volume, equal shape, not oriented;
- EEI – diagonal, equal volume, equal shape, axes centrally symmetric;
- VEI – diagonal, varying volume, equal shape, axes centrally symmetric;
- EVI – diagonal, equal volume, varying shape, axes centrally symmetric;
- VVI – diagonal, varying volume, varying shape, axes centrally symmetric;
- EEE – ellipsoidal, equal volume, shape and orientation;

- EEV – ellipsoidal, equal volume and equal shape;
- VEV – ellipsoidal, equal shape;
- VVV – ellipsoidal, varying volume, shape and orientation.

The Figure 1 presents results of the *mclust* procedure employed for the estimation of the optimal number of components in the hierarchical clustering of random effects in linear mixture model with autoregressive component (for detail see Forbelská and Bartošová, 2010). Maximal value of the Bayesian information criterion (*BIC* = 16390) was achieved for the four-component mixture with covariance structure VVV.

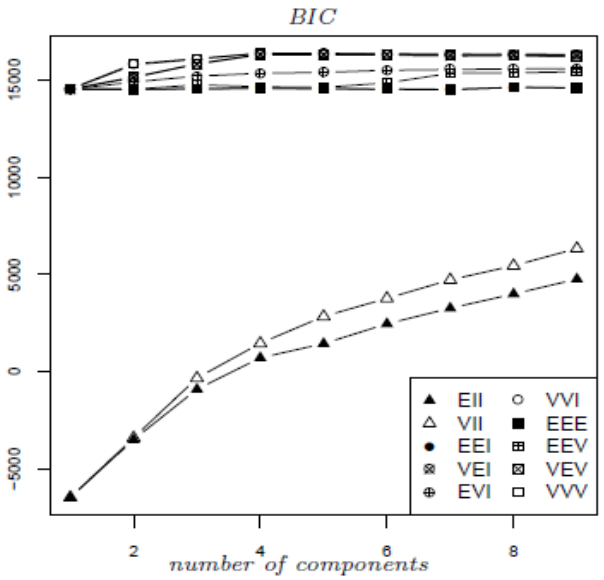


Figure 1 Values of Schwarz's Bayesian information criterion (*BIC*) for the estimation of the optimal number of components for different triplets E, V, I.

[Source: R code output (Forbelská and Bartošová, 2010)]

2 Methods for component agglomeration

But the number of clusters based on the optimization of information criterion quite often does not agree with reality. The estimated number is usually higher than the real one. Such example is presented, e.g., in paper by Hennig (2010) where the optimal number of clusters found using BIC criterion was 9 but there were only 4 real clusters. So it is necessary to search for methods capable to merge the clusters into larger aggregates. There exist several different approaches, e.g., Tantrum, Murua and Stuetzle (2003) propose graphical diagnostic methods and recommend to merge the clusters of normal mixtures using hierarchical methods based on the dip test statistic designed for the testing of unimodality of distributions (see Hartigan and Hartigan, 1985). Also non-hierarchical methods exist, e.g., based on the normal mixtures, some of them apply for the agglomeration well-known k-means method, etc.

In normal mixtures the methods of cluster agglomeration into the form of (separable) clusters are described also in the paper of (2010) who checked several different methods, but many of the with objections. Therefore, he proposed other concepts for cluster agglomeration different from the unimodality-based methods. He designed the DEMP method based on the direct estimate of probability of incorrect classification (for details, see again Hennig, 2010). There exists an explicit formula for determination of distance in case of two multidimensional distributions (Fukunaga, 1990). It is based on the determinant and matrix inverse to the covariance matrix of the components. Its value cannot be accurately estimated and the results can be (in case of nearly singular matrices) significantly biased. Huge amount of papers concern also the methods for determination of the number of clusters.

2.1 Separation and confusion

The clustering of mixture models cannot achieve perfect separation of objects or variables. We can only obtain information concerning the probability of membership of the object in the given cluster. The resulting clusters will be almost always more or less overlapping. The question arises, to what extent can be the particular components considered as clusters – to what extent is the assignment of the objects into the clusters "objective". This is closely related to the extent of separation of the clusters – whether or not some subset or subpopulation

forms a cluster. Let us state a question: "Can be all components considered as clusters or is it only case of come (original or agglomerated) components, respectively the sets of agglomerated components of the mixture model?" Component will be considered as cluster only in case of "sufficient" separation from other components (clusters). In such case the probability of assignment of an observation to the "correct" cluster is high, although we are not sure about the assignment to a cluster (clusters are latent). The above described method extends the Hennig's concept based on the direct estimation of probability of wrong classification (DEMP method, see Hennig, 2010).

Any component of the mixture (latent cluster) is in our case described by a pair (p, D), where p is a marginal probability of the component (weight of the cluster) and D is a distribution of the component given by its density f, resp. scaled density $pf(x)$. The component can be considered as a cluster only with respect to some other component. In case of working with the group of two or more components we can observe cluster

- individually – any component is a cluster with respect to the union of remaining components in the given set;
- pairwise – every component is a cluster with respect to any other component.

For the pairs of clusters the crucial information is a confusion index. For components $A = (p_A, D_A)$ and $B = (p_B, D_B)$ the confusion index of A and B is given by probability $r_{A|B}$, that to the component A will be assigned a random sample from the distribution D_B. The assignment process of the objects to the components in EM algorithm is driven by the Bayes theorem, p_A and p_B are the probabilities of events A, resp. B. Two components can be considered as cluster if they are sufficiently separated, i.e., if the probabilities of their confusion $r_{A|B}$ and $r_{B|A}$ are smaller than some threshold T chosen in advance.

Classification rule is (like in E-step of the EM algorithm – see Dempster, Laird and Rubin, 1977) given by the conditional probability that the observation x from the domain of densities f_A, resp. f_B is from the component A, resp. B, i.e., value

$$r_{x,A|AB} = \frac{p_A f_A(x)}{p_A f_A(x) + p_B f_B(x)} \tag{3}$$

resp. value $r_{x,B|AB} = 1 - r_{x,A|AB}$. The observation x thus belongs to A with probability $r_{x,A|AB}$ and to B with probability $1 - r_{x,A|AB}$.[2] Particular assignments of the sets of objects are mutually independent. The index of confusion of A and B is then given by the expected probability that the component A has assigned some random sample from distribution D_B, i.e.,

$$r_{A|B} = \int \frac{p_A f_A(x) f_B(x)}{p_A f_A(x) + p_B f_B(x)} dx \tag{4}$$

Analytical expression of the index of confusion does not exist for any nontrivial distribution (with the exception of a pair of uniform distributions). But we can approximate it with arbitrary precision by its empirical variant if we get a large random sample from B and average the probabilities in the formula for $r_{x,A|AB}$. For an arbitrary pair p_A a p_B the index of confusion depends only on the probability ratio p_B / p_A and for $p_A = p_B$ it is $r_{A|B} = r_{B|A}$. For the mixture of k components $A_1, \ldots, A_k$ we can construct a confusion matrix

$$\mathbf{R} = \begin{pmatrix} r_{1|1} & \cdots & r_{1|k} \\ \cdots & r_{i|j} & \cdots \\ r_{k|1} & \cdots & r_{k|k} \end{pmatrix} \tag{5}$$

where $r_{i|j}$ represents the index of confusion of the components A_i and A_j.

3 Application

The measured separation, resp. confusion of the components can be illustrated either using simulation or by an application to the real data files.

[2] Simultaneously we assume that x actually comes either from A or B, but not from both.

3.1 Mixture simulation

The situation is documented on the Figure 2, where the scaled two and three (normally distributed) component mixture is shown. In the three component case, the densities $p_1 f_1$, $p_2 f_2$ and $p_3 f_3$ are depicted by thin lines and the density of their mixture $p_1 f_1 + p_2 f_2 + p_3 f_3$ is depicted by a thick line. Left part of the figure presents simplified situation for two components. The picture presents that the particular marginal densities overlap significantly and therefore, we can anticipate than neither of the components can be considered as cluster, in spite of the fact that two components with lower variability have substantial distance.

The decision made in case of particular pairs of components – whether we will consider them as

- sufficiently separated;
- mutually confusable;
- satellite – one of them is a subset of the other;

depends on the level of the threshold T. The estimate of the cluster structure, i.e., ascertaining of the manner how the components are agglomerated in the model, to some extent depends on the choice of threshold value.

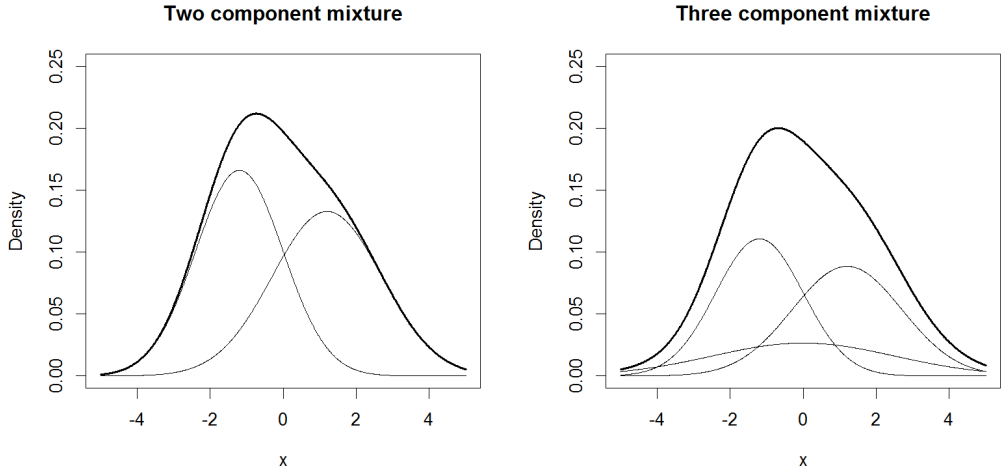


Figure 2 Simulation of the mixture of two, resp. three normal components.

3.2 Real mixture model

Application of the index of confusion is also demonstrated on the real data files. For this purpose the cross-sectional data from sample survey EU-SILC 2010 were used, namely the equalized values of net annual incomes from year 2009. According to the large scope of the national panel of EU-SILC surveys the penalized criteria (*AIC*, resp. *BIC*) often suggest use of high amount of components. But these components in the vast majority have very small marginal probabilities and therefore the decomposition is inappropriate. From the up to now performed analyses we found out that for some countries the four-component mixture ($K = 4$) is appropriate and for most of the countries the best model is a mixture of five components ($K = 5$). The singularity (i.e., a probability p_k converging to zero) can be found in case of only few countries where the best model as an output of the *mclust* procedure is a six-component mixture.

The Figure 3 depicts a kernel estimate of the empirical distribution of the logarithms of equalized household incomes in case of Czech household (dashed line) and the corresponding mixture of four normal components (solid lines). The estimation of parameters of mixture model was based on the *mclust* procedure implemented R in software. The components $A_1 = (p_1, D_1)$, $A_2 = (p_2, D_2)$, $A_3 = (p_3, D_3)$ and $A_4 = (p_4, D_4)$ are computed, where $D_1 = N(12.127;\ 0.142)$, $D_2 = N(11.872;\ 0.088)$, $D_3 = N(12.012;\ 0.411)$ and $D_4 = N(11.444;\ 1.649)$ and $p_1 = 0.554$, $p_2 = 0.340$, $p_3 = 0.097$ and $p_4 = 0.009$. The confusion matrix of the four components (A_1, A_2, A_3, A_4) is following

$$\mathbf{R} = \begin{pmatrix} 1 & 0.123 & 0.007 & 0.000 \\ 0.224 & 1 & 0.030 & 0.000 \\ 0.037 & 0.258 & 1 & 0.004 \\ 0.000 & 0.003 & 0.027 & 1 \end{pmatrix}. \tag{6}$$

The choice of *threshold* $T = 0{,}01$ leads to a component A_4 sufficiently separated from the components A_1 and A_2 and is a satellite of the component A_3. In contrast, components A_1 and A_2 will be confusable similarly like components A_2 and A_3. If we choose $T = 0{,}05$, the components A_1 and A_3 will be sufficiently separated and component A_3 will be a satellite of component A_4. Component A_4 will be sufficiently separated from other components and will form a cluster.

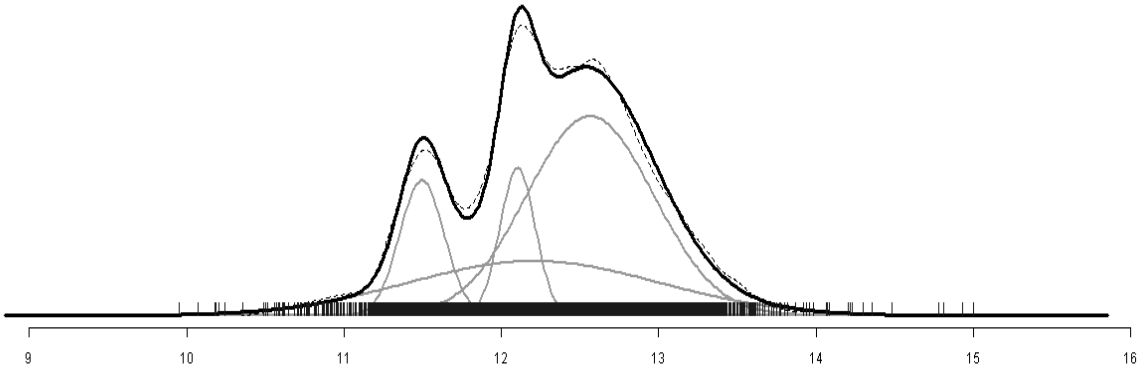


Figure 3 Kernel estimate of the empirical distribution of logarithms in equalized annual incomes (dashed line) and corresponding model of mixture of four normal components (solid lines) (EU-SILC 2010 – cross-sectional data).

4 Conclusion

The presented paper addresses the issues of large dataset modelling and concerns the approach of finite mixture models. It has been shown that components found in the data are not necessarily sufficiently separated and therefore can be under certain conditions again agglomerated. As a measure of separation authors present index of confusion. Values greater than some suitable threshold indicate low degree of separation (satellite and mutually confusable components).

Acknowledgements

This research was supported by internal project of the University of Economics in Prague (no. F6/30/2014).

References

[1] Akaike, H.: On entropy maximization principle. In: *Applications of Statistics* (Krishnaiah, P.R., ed.), North-Holland, Amsterdam, 1977, 27–41.

[2] Bartošová, J.: *Finanční potenciál domácností – kvantitativní metody a analýzy*. Professional Publishing, Praha, 2010.

[3] Dempster, A., Laird, N., and Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39,** 1 (1977), 1–38.

[4] Forbelská, M., and Bartošová, J.: Clustering of Czech Household Incomes Over Very Short Time Period. Paris In: *Proceedings of COMPSTAT 2010*. Physica-Verlag HD, Paris, 2010, 1015 – 1022.

[5] Fraley, C., and Raftery, A. E.: Enhanced Software for Model-Based Clustering: Density Estimation and Discriminant Analysis. *Journal od Classification*, **20** (2003), 263–286.

[6] Fukunaga, K.: *Introduction to statistical pattern recognition*, Academic Press, Boston, 1990.

[7] Hartigan, J. A., and Hartigan, P. M.: The dip test of unimodality. *Annals of Statistics*, **13**(1985), 70–84.

[8] Hennig, C. H.: Methods for merging Gaussian mixture components. *Advances in Data Analysis and Classification* **4** (2010), 3–34.

[9] Řezanková, H., Húsek, R., and Snášel, V.: *Shluková analýza dat*. Professional Publishing, Praha, 2009.

[10] Schwarz, G. E.: Estimating the dimension of a model. *The Annals of Statistics*, **6**, 2 (1978), 461–464.

[11] Tantrum, J., Murua, A., and Stuetzle, W.: Assessment and Pruning of Hierarchical Model Based Clustering. In: *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*. Washington, D.C., 2003, 197–205.

Statistical inference about modality of random variable

Aleksandra Baszczyńska[1], Dorota Pekasiewicz [2]

Abstract. In the paper the detailed analysis of a random variable distribution, especially the modality, is considered. In a number of disciplines, including economy, this issue is treated as one of the most considerable interest. Detecting multimodality in income distributions, as well as the analysis of economic polarization are the examples of application of multimodality analysis in explaining economic phenomena. The paper presents statistical methods of detecting the multimodality. The bimodality coefficients, the biaverage and other methods of statistical inference are considered. The bimodality coefficient allows to distinguish between unimodality and bimodality of the random variable distribution. Furthermore, statistical tests (Silverman test, Hartigan Dip test) can be used in verification of hypothesis about unimodality. Chosen estimators of the mode in case of modality are used. For the bimodal distribution, biaverage is estimated. Apart from theoretical considerations, the results of simulation study on the properties of these methods are presented. The results of the study indicate the shortcomings and advantages of the considered methods.

Keywords: biaverage, bimodal coefficient, modality, Hartigan Dip test.

JEL Classification: C12, C13, C15
AMS Classification: 62G05, 62G10

1 Introduction

Assessing modality of a distribution of a random variable is one of the most important, but still not sufficiently investigated issues, both from theoretical and practical point of view. Historical overview of the problem of modality is presented in [9]. In a number of disciplines there have been some attempts to use this characteristic of a random variable in statistical analysis, presented for example in [3], [5], [6], [7] and [10]. But a strong need can be observed for further analysis in the area of detecting modality and identification modes, as well as indicating possibilities of applications, in economy, biology or in technical sciences.

In the paper, the definition of the mode in one dimension, commonly used in the statistical literature, is applied. The density f is unimodal with the mode m, if there exists a real number m such that f is non-decreasing on $(-\infty,m)$ and non-increasing on (m,∞). In other cases, density f is multimodal, which can we interpreted as a mixture distribution, containing several subpopulations. So, a mode in a density is a local maximum, while an interval $[a,b]$ such that density f is concave only on $[a,b]$ is defined as a bump. According to this definition uniform distribution is a unimodal one.

Different methods of assessing modality can be used, aimed at different goals. In particular, in the analysis of income distribution or in the analysis of economic polarization, the following set of modality techniques is proposed: calculating the bimodality coefficient, testing the unimodality and in case of the bimodal distribution calculating the biaverage as the estimator of modes.

In the paper chosen modality procedures are presented, paying particular regard to their range of applications.

2 Assessing modality

The problem of verification hypotheses about the number of modes of the random variable *X* can be formulated in the following way:

[1] University of Lodz/Department of Statistical Methods, 90-214 Lodz, 41/43 Rewolucji 1905 St. Poland, albasz@uni.lodz.pl.
[2] University of Lodz/Department of Statistical Methods, 90-214 Lodz, 41/43 Rewolucji 1905 St. Poland, pekasiewicz@uni.lodz.pl.

$$\begin{cases} H_0: \ j=1, \\ H_1: \ j>1, \end{cases} \tag{1}$$

where j is the number of modes in the considered population. The hypothesis H_0 means that the distribution of X is unimodal while the alternative hypothesis states that it is multimodal. There are some statistical tests that can be applied to test the unimodality.

Let $X_1, X_2, ..., X_n$ be the random sample of size n from the continuous random variable X with the density f and the distribution function F. The Hartigan Dip test and the Silverman test are examples of the most widely used tests in assessing the unimodality.

The Hartigan Dip test

The test statistic is based on the dip $D(F)$, which is defined as:

$$D(F) \equiv \delta(F,U), \tag{2}$$

where $\delta(F,U) \equiv \inf_{G \in U} \sup_x |F(x) - G(x)|$ and U is the class of unimodal cumulative distribution functions. Then $D(F)$ is the value that minimizes this maximum difference between the distribution function F and the unimodal distribution function. It means that the dip measures the departure from unimodality.

Glivenko-Cantelli theorem states that distance $\delta(F_n, F) \to 0$ a.s., where F_n is the empirical distribution, and so $D(F_n) \to D(F)$ a.s. The details of the numerical method of computing $D(F_n)$ is presented in [8]. The probability of the rejection of the unimodal distribution is calculated empirically and tabulated as a function of the sample size.

The Silverman test

Verification of the hypothesis about unimodality of the distribution of a random variable can be done through the bootstrap method. Silverman bootstrap test is the example of a statistical test of this kind. We estimate function f by the kernel density estimator with Gaussian kernel function (cf. [11]), which is defined in the following way:

$$\hat{f}(x,h) = \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - x_i}{h} \right)^2 \right), \tag{3}$$

where: n is the sample size, $x_1, x_2, ..., x_n$ is the realization of $X_1, X_2, ..., X_n$, h is the smoothing parameter. The test statistic (cf. [4]) is the following:

$$t = \min_{h>0} \left\{ h: \quad \hat{f}(x,h) \ \text{ is unimodal} \right\}. \tag{4}$$

We generate N, for example N=10000, bootstrap samples $x_1^*, x_2^*, ..., x_n^*$ and the values of the standardized normally distributed random variable $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$. Next we create:

$$y_i^* = x_i^* + h\varepsilon_i \quad \text{for} \quad i = 1, ..., n. \tag{5}$$

For each sample $y_1^*, y_2^*, ..., y_n^*$ we estimate the density function f and calculate t^* using the formula (4). After N-th repetitions we obtain $t_1^*, t_2^*, ..., t_N^*$ statistics and we can calculate the *p-value* in the following way:

$$p = \frac{\#\left\{ t^*: \ t^* \geq t \right\}}{N} \tag{4}$$

Bimodality coefficient

In assessing bimodality the coefficient BC can be used, defined as:

$$BC = \frac{m_3^2 + 1}{m_4 + 3 \cdot \frac{(n-1)^2}{(n-2)(n-3)}}, \tag{6}$$

where m_3 is the skewness of the empirical distribution and m_4 is its excess kurtosis.

The *BC* of a given empirical distribution is compared to the benchmark value of $BC_{crit} = \frac{5}{9} \approx 0{,}555$, which indicates the uniform distribution. Higher values than 0,555 show the bimodality while lower values show unimodality of the considered random variable distribution. The maximum value of the bimodality coefficient (*BC*=1) is obtained for the Bernoulli distribution (cf. [10]). It means that the number of modes influences the value of the coefficient but in case of a distribution with heavy-tails the coefficient BC has smaller values regardless of the number of modes.

3 Biaverage for bimodal distribution

The term biaverage for the bimodal distribution and, as generalization, *k*-average for the *k*-modal distribution was proposed by Antoniewicz in [1]. Biaverage is defined as a pair of parameters given by the concentration of probability.

When the bimodal distributed random variable *X* has four moments and its variance is not equal to zero, then the biaverage (m_1, m_2) is the following:

$$m_1 = \frac{1}{2}\left(P - \sqrt{P^2 + 4Q} \right), \tag{7}$$

$$m_2 = \frac{1}{2}\left(P + \sqrt{P^2 + 4Q} \right), \tag{8}$$

where $P = \frac{E\left(X^3\right) - E\left(X^2\right)E(X)}{E\left(X^2\right) - E^2(X)}$ and $Q = \frac{E^2\left(X^2\right) - E\left(X^3\right)E(X)}{E\left(X^2\right) - E^2(X)}$.

The values of the biaverage can be estimated from the random sample $X_1, X_2, ..., X_n$, drawn from a bimodal population, through the following formulas:

$$\hat{m}_1 = \frac{1}{2}\left(p - \sqrt{p^2 + 4q} \right), \tag{9}$$

$$\hat{m}_2 = \frac{1}{2}\left(p + \sqrt{p^2 + 4q} \right), \tag{10}$$

where $p = \frac{n\sum_{i=1}^{n} x_i^3 - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$ and $q = \frac{\left(\sum_{i=1}^{n} x_i^2\right)^2 - \sum_{i=1}^{n} x_i^3 \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$.

It is shown in [2] that if sample moments are good estimators of population moments, the biaverage is a good estimator of modes (small bias and variance) and specifies concentration of probability. For a random variable with heavy-tails distribution some moments do not exist, so we cannot estimate the modes in this way.

4 Simulation study

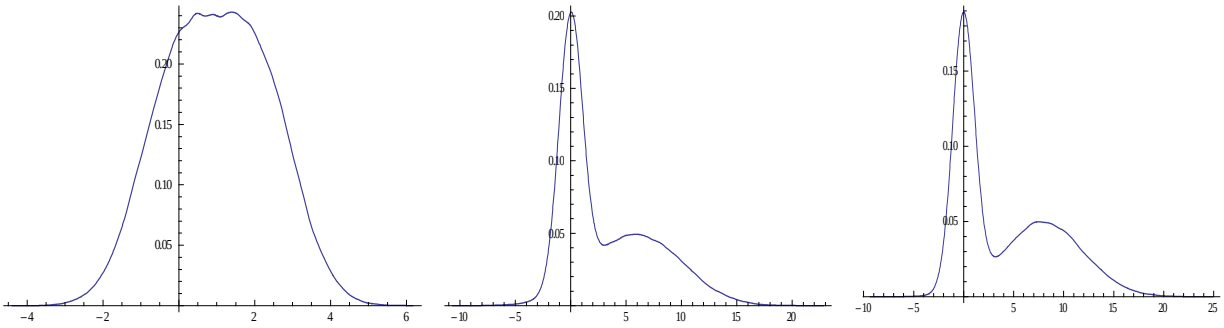
The simulation study was conducted to investigate the properties of chosen modality methods described above. In the study, the Hartigan Dip test, the bimodality coefficient and the biaverage were used and the results of applying these methods were analyzed and compared. Chosen modality methods were applied to samples drawn from populations which are mixtures of normal distributions and normal and exponential distributions. The density $f(x)$ can be written as a two-component mixture $f(x) = wf_1(x) + (1-w)f_2(x)$, where $w \in (0,1)$ and f_1, f_2 are known as component densities of $f(x)$ in a mixture of two distributions. In the study, different values of w and different samples sizes *n* were used. The choice of component densities was based on our earlier studies as well as it was motivated by the attempt to present unimodal distributions and multimodal ones.

Table 1 presents results of applying the bimodality coefficient and *p*-values of the Hartigan Dip test for chosen mixtures of distributions and chosen sample sizes. The value of w for the mixture of population distribution was set to 0.5.

Components of mixture	n	BC	p-value for Hartigan Dip test
$N(0,1)$ and $N(2,1)$	100	0.4532	0.9917
	300	0.4289	0.6236
	500	0.4100	0.9864
$N(0,1)$ and $N(6,4)$	100	0.6183	0.8814
	300	0.6433	0.9983
	500	0.6220	0.8678
$N(0,1)$ and $N(8,4)$	100	0.6717	0.5084
	300	0.5952	0.7418
	500	0.6307	0.0194
$Exp(2)$ and $N(2,1)$	100	0.5354	0.0312
	300	0.6145	0.8432
	500	0.5682	0.9891
$Exp(2)$ and $N(6,4)$	100	0.6511	0.5464
	300	0.6457	0.3204
	500	0.6444	0.9256
$Exp(2)$ and $N(8,4)$	100	0.6400	0.9941
	300	0.6129	0.4466
	500	0.6569	0.0034

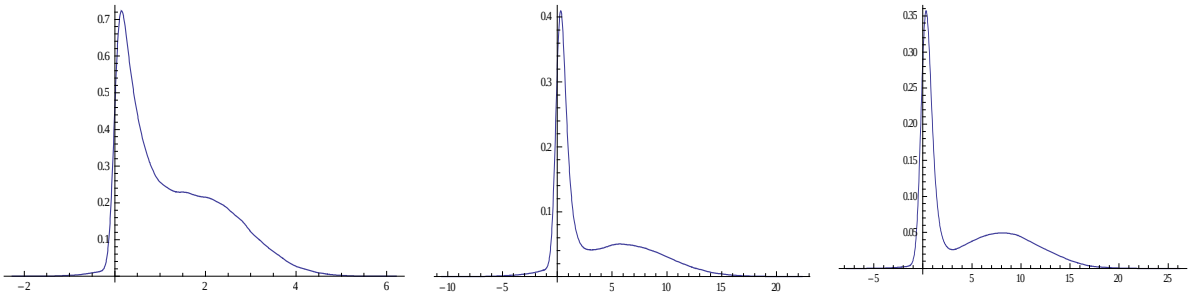
Table 1 *BC* coefficients and *p*-values for the Hartigan Dip test for mixtures of distributions

It is worth noticing that in case of the unimodal distribution of the population (for example population with components $N(0,1)$ and $N(2,1)$) the bimodality coefficient as well as the Hartigan Dip test indicate the unimodality. By contrast, for the multimodal distribution of the population, for example the population with components $N(0,1)$ and $N(8,4)$ and the population with components $Exp(2)$ and $N(6,4)$, in case of n=300, *BC* indicates bimodality whereas the Hartigan Dip test – the unimodal distribution. The obtained values of *BC* for mixtures of normal distributions and mixtures of normal and exponential distributions are consistent with theoretical distributions of populations. Figure 1 and 2 present distributions of chosen regarded populations.



A. $N(0,1)$, $N(2,1)$ *B.* $N(0,1)$, $N(6,4)$ *C.* $N(0,1)$, $N(8,4)$

Figure 1 Distributions for mixtures of normal distributions for w=0.5



A. $Exp(2)$, $N(2,1)$ *B.* $Exp(2)$, $N(6,4)$ *C.* $Exp(2)$, $N(8,4)$

Figure 2 Distributions for mixtures of exponential and normal distributions

In the next stage of simulation study, different mixtures of distributions were used for chosen values of w. The bimodality coefficient and the Hartigan Dip test were calculated for samples drawn from the populations. The results for mixture $N(0,1)$ and $N(8,4)$ are presented in Table 2.

n	Methods	w							
		0.1	0.2	0.3	0.4	0.6	0.7	0.8	0.9
100	*BC*	0.4329	0.4956	0.4986	0.5259	0.7040	0.7537	0.7402	0.8321
	Hartigan test (*p*-value)	0.8385	0.0992	0.2392	0.3355	0.5781	0.9626	0.9523	0.6953
300	*BC*	0.4318	0.4648	0.5036	0.5970	0.6864	0.7343	0.7711	0.7984
	Hartigan test (*p*-value)	0.8341	0.0033	0.00008	0.0811	0.7948	0.7659	0.9917	0.9924
500	*BC*	0.4011	0.4430	0.5265	0.6008	0.6861	0.7617	0.7598	0.7405
	Hartigan test (*p*-value)	0.7943	0.0009	0.00002	0.00005	0.1693	0.3155	0.9662	0.0999

Table 2 *BC* coefficients and *p*-values for Hartigan Dip test for mixtures of distributions with components $N(0,1)$ and $N(8,4)$ and chosen w

Taking into account the bimodality coefficient *BC*, the bimodal populations were indicated. The values of the biaverage for such distributions are presented in Table 3.

n	*biaverage*	w					
		0.4	0.5	0.6	0.7	0.8	0.9
100	$\hat{m}_1$	0.9130	0.3355	0.1661	0.1850	-0.1703	-0.1477
	$\hat{m}_2$	9.9152	11.3546	10.7135	10.867	9.4329	9.3938
300	$\hat{m}_1$	0.7504	0.8070	0.4200	0.1338	-0.0452	-0.0134
	$\hat{m}_2$	10.6711	11.4064	11.1991	10.9067	9.8411	9.9206
500	$\hat{m}_1$	0.8207	0.4504	0.4685	0.0136	-0.0578	-0.1227
	$\hat{m}_2$	10.8358	11.1314	11.5078	9.8655	10.2562	9.9451

Table 3 Biaverages for mixtures of distributions with components $N(0,1)$ and $N(8,4)$ and chosen w

The results presented in tables 1-2 are only the illustration of the applications of considered methods. The generalization of these results need deeper analysis, including simulation methods. Properties of the bimodal coefficient *BC* were analyzed through simulation methods. The samples were drawn from the previously considered populations and *BC* were calculated. The number of repetitions was set to 50 000. The results are presented in Table 4. Obtained results for w=0,5 concerned unimodality or bimodality of distribution are consistent with information about modality of population, from which samples were drawn.

Components of mixture	n	Min *BC*	Max *BC*	Mean *BC*	Fraction of rejection H_0
$N(0,1)$ and $N(2,1)$	100	0.2246	0.5654	0.4051	0.0000
	300	0.2970	0.5066	0.4040	0.0000
	500	0.3165	0.4901	0.4018	0.0000
$N(0,1)$ and $N(6,4)$	100	0.1656	0.7740	0.6076	0.8482
	300	0.4332	0.7370	0.6105	0.9336
	500	0.4598	0.7110	0.6100	0.9622
$N(0,1)$ and $N(8,4)$	100	0.3976	0.7927	0.6260	0.9303
	300	0.5071	0.7507	0.6373	0.9956
	500	0.5091	0.7214	0.6370	0.9992

Exp(2) and *N*(2,1)	100	0.2995	0.7519	0.5685	0.6283
	300	0.3705	0.6876	0.5728	0.7313
	500	0.4255	0.6637	0.5752	0.7937
Exp(2) and *N*(6,4)	100	0.1707	0.8137	0.6374	0.8919
	300	0.3736	0.7677	0.6400	0.9613
	500	0.4606	0.7461	0.6394	0.9845
Exp(2) and *N*(8,4)	100	0.4136	0.8195	0.6603	0.9746
	300	0.5091	0.7826	0.6631	0.9977
	500	0.5528	0.7526	0.6671	0.9999

Table 4 Chosen characteristics of *BC* coefficient distribution and fraction of rejection hypothesis about unimodality

5 Conclusions

In the paper chosen methods of analyzing modality are presented. The simulation study showed that for the mixture of two normal distributions and the mixture of the exponential and the normal ones, the bimodal coefficient has good properties. In almost all cases, especially for big samples, this coefficient properly indicates bimodality. These results were confirmed by simulation methods. For the Hartigan Dip test the calculated *p*-value didn't allow to reject the hypothesis about unimodality in most cases, which indicates the need for further analysis of this method.

The advantage of *BC* coefficient is the simplicity of computing it (only values of skewness and kurtosis are needed). The results of simulation analysis for the mixtures of distributions, presented above, indicate that there is a possibility of using this measure in assessing the characteristics of random variables which are the mixtures of distributions.

The Hartigan Dip test seems to be more complicated and sophisticated but it is well-programmed (for example in R and Matlab) what makes it more useful in practical applications. In case of regarded mixtures of distributions it indicated unimodality even when population was bimodal.

For bimodal distributions the biaverage seems to be useful, because it gives values of estimated parameters connected with the concentration of probability of the random variable.

References

[1] Antoniewicz, R.: O średnich i przeciętnych. *Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu*, Wrocław, 2005.
[2] Antoniewicz, R.: Pięć lat pojęcia dwuprzeciętnej. *Zeszyty Naukowe Uniwersytetu Szczecińskiego* **450** (2006), Szczecin, 11-15.
[3] Baszczyńska, A., Pekasiewicz, D.: Biavarage and Multimodality in Investigating Distribution of Electricity Prices, *Folia Oeconomica, Acta Universitatis Lodziensis* **302** (2014), Łódź.
[4] Davison, A.C., Hinkley, D.V.: Bootstrap Methods and their Application. *Cambrigde University Press*, 1997.
[5] Freeman, J., Dale, R.: Assessing Bimodality to Detect the Presence of a Dual Cognitive Process. *Behav. Res. Methods* **45** (2013), *83-97*, doi: 10.3758/s13428-012-0225-x.
[6] Henderson, D., Attar, P., Russell R.*:* Modality Tests for Use in Applied Econometrics. *Application to Macroeconomic Convergence*, 2002, unpublished paper.
[7] Henderson, D., Parmeter, C., Russell R.: Modes, Weighted Modes, and Calibrated Modes: Evidence of Clustring Using Modality Tests. *Journal of Applied Econometrics* **23** (2008), 607-638.
[8] Hartigan, J., Hartigan, P.: The Dip Test of Unimodality. *The Annals of Statistics* **13** (1985), 1, 70-84.
[9] Knapp, T.: Bimodality Revisited. *Journal of Modern Applied Statistical Methods* **6** (2007), 8-19.
[10] Pfister, R., Schwarz, K., Janczyk, M., Dale, R., Freeman, J.: Good Things Peak in Pairs: a Note on the Bimodality Coefficient, *Frontiers in Psychology* **4** (2013), doi: 10.3389/fpsyg.2013.00700.
[11] Silverman, B. W.: Using Kernel Density Estimates to Investigate Multimodality. *Journal of the Royal Statistical Society, Series B (Methodological)* **43 (**1981), No. 1, 97-99.

Social choice function and multiple criteria group decision making

Iveta Bebčáková[1], Pavla Kouřilová[2]

Abstract. The multiple-criteria group decision making problem is quite difficult to solve. It consists of two partial problems: Multiple criteria decision making problem and Group decision making problem. Both of the problems have been studied quite intensively and many methods and approaches were introduced (for more details see [1], [2], [3], [4], [5], [6]).

In the paper, we focus on the process of combining the results of those partial problems with help of two social choice functions: Borda function and Copeland function. We apply two different approaches of finding a solution and discuss the results.

Keywords: social choice function, Borda function, Copeland function, MCDM, group decision making, aggregation.

JEL classification: C44
AMS classification: 90B50

1 Introduction

Every day of our lives we can meet a multiple criteria group decision making problem (project selection, university admission, beauty contest, chefs competition, wine competition, etc.). Such a problem can be described as follows. There is a set of alternatives (contestants, applicants, wines, chefs) which are evaluated with respect to various criteria by several experts. The goal is always the same: to order the alternatives and find out which one is the best one. Sometimes we need to know how good the best alternative is, e.g. some project could be the best compared to the other ones, but is it good enough to invest in? In other situations the absolute evaluation of particular alternative is not important and the ordering of the alternatives is the only solution we want to find. In this paper we focus just on ordering of the alternatives.

2 Notation and methods

Let $A = \{A_1, \ldots, A_n\}$ be a set of alternatives which are supposed to be ordered from the best one to the worst one. Let $E = \{E_1, \ldots, E_r\}$ be a set of experts and $C = \{C_1, \ldots, C_m\}$ be a set of criteria. Let us assume that the criteria are of the same importance and the experts are of the same level of expertise. Therefore, we do not take weights of criteria or of experts into account. We assume that each expert E_j, $j = 1, \ldots, r$, is able to order the alternatives with respect to each criterion C_i, $i = 1, \ldots m$, i.e. we have $r \cdot m$ initial orderings of the alternatives. Moreover, we assume that there can be some alternatives which share the same place in the initial orderings. Let us denote by $\mathcal{O}_{ji}$ the initial ordering of the alternatives $A_1, \ldots A_n$ stated by the expert E_j with respect to the criterion C_i, for all i, j.

To aggregate these initial orderings into the overall one we can apply two approaches (Approach I and Approach II). Both of them consist of two steps of aggregation.

In Approach I (see Figure 1 left), we start with the experts E_j, $j = 1, \ldots, r$. For each $j \in \{1, \ldots r\}$ we aggregate the initial orderings $\mathcal{O}_{j1}, \mathcal{O}_{j2}, \ldots \mathcal{O}_{jm}$ into one ordering denoted by $\mathcal{O}_{j\cdot}$, i.e. into the ordering stated by the expert E_j with respect to all criteria together. These first-level orderings $\mathcal{O}_{1\cdot}, \mathcal{O}_{2\cdot}, \ldots \mathcal{O}_{r\cdot}$

[1]Palacký University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, iveta.bebcakova@upol.cz
[2]Palacký University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, pavla.kourilova@upol.cz

are then again aggregated in the second step to obtain the final ordering $\overline{\mathcal{O}} := \mathcal{O}_{..}$ of the alternatives w.r.t. all experts and criteria.

Approach II works similarly, but the aggregations are made in reverse order (Figure 1 right). At first we focus on the criteria. For each $i \in \{1, \dots m\}$ we aggregate the initial orderings $\mathcal{O}_{1i}, \mathcal{O}_{2i}, \dots \mathcal{O}_{ri}$ into one ordering denoted by $\mathcal{O}_{\cdot i}$, i.e. into the ordering stated with respect to the criterion C_i by all experts together. These first-level orderings $\mathcal{O}_{\cdot 1}, \mathcal{O}_{\cdot 2}, \dots \mathcal{O}_{\cdot m}$ are then again aggregated in the second step to obtain the final ordering $\hat{\mathcal{O}} := \mathcal{O}_{..}$ of the alternatives w.r.t. all experts and criteria.

Since both approaches are used in various competitions, some questions emerged: Are these two final orderings $\overline{\mathcal{O}}$ and $\hat{\mathcal{O}}$ always the same or does the final ordering depend on the order of aggregations? What kind of alternatives do particular approaches offer as a winner?

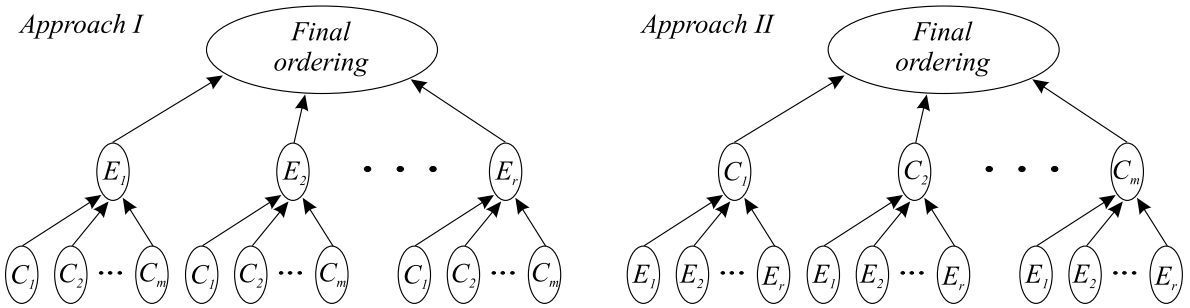


Figure 1 Two levels of aggregation: Approach I and Approach II.

Example 1. As an example we can consider a beauty contest. The contestants are the alternatives which have to be ordered from the best one to the worst one, the criteria correspond to particular competition tasks. Applying Approach I, all contestants face various kinds of challenges and then each judge makes a decision about the ordering of contestants based on their performance through the whole competition. Particular orderings of the judges are then aggregated into the overall ordering of the contestants. Using Approach II, the overall ordering is found in other way. After each challenge, the judges make a group decision about the ordering of the contestants (based on their individual preferences). Then, after all the challenges, these first-level orderings (i.e. results of each discipline) are aggregated and the final ordering is announced.

Whether we choose the Approach I or Approach II, the process of aggregation is crucial and can be done in many ways ([4],[5], [6]). In this paper we will apply the social choice functions ([1],[3]). Usually, social choice functions are associated with election systems. They were designed to deal with a problem of preferential elections, i.e. the problem of searching for right ordering of candidates based on preferences of the voters. Nevertheless, as the criteria can play the role of voters, the social choice functions can be applied also to multiple criteria decision making problem.

In the rest of the paper we will use the following notation:

- $\mathrm{Card}(A) \dots$ the cardinality of the set A
- $A_p P_j^i A_q \dots$ the alternative A_p is preferred to the alternative A_q by expert E_j with respect to criterion C_i.
- $A_p P_j A_q \dots$ the alternative A_p is preferred to the alternative A_q by expert E_j (after all the criteria were considered).
- $A_p P^i A_q \dots$ the alternative A_p is preferred to the alternative A_q with respect to criterion C_i (after all the experts' opinions were taken into account).
- $A_p P A_q \dots$ the alternative A_p is preferred to the alternative A_q (the overall preference).
- $A_p I A_q \dots$ the alternatives A_p and A_q are indifferent
- $M_{pq}^i = \{E_j; A_p P_j^i A_q\} \dots$ the set of experts which prefer the alternative A_p to the alternative A_q with respect to criterion C_i

- $N_{pq}^j = \{C_i; A_p P_j^i A_q\} \dots$ the set of criteria according to which the alternative A_p is preferred to the alternative A_q by expert E_j
- $M_{pq} = \{E_j; A_p P_j A_q\} \dots$ the set of experts which prefer the alternative A_p to the alternative A_q (after all the criteria were considered)
- $N_{pq} = \{C_i; A_p P^i A_q\} \dots$ the set of criteria according to which the alternative A_p is preferred to the alternative A_q (after all the experts' opinions were taken into account)

Both of the approaches mentioned above are based on the assumption that each expert E_j, $j = 1, \dots, r$, is able to order the alternatives with respect to each criterion C_i, $i = 1, \dots, m$. Therefore, for each pair of alternatives A_p, A_q, $p, q = 1, \dots, n$ the expert E_j, $j = 1, \dots, r$, is able to state if $A_p P_j^i A_q$ or $A_q P_j^i A_p$ or both alternatives are indifferent w.r.t criterion C_i, $i = 1, \dots, m$.

We will use the modifications of two social choice functions: Borda function and Copeland function ([1],[3]). The principle of two-level aggregation will be explained with help of Borda function. Copeland function is applied analogically.

The process of aggregation according to Approach I with help of modified Borda function can be described as follows:

1. First-level aggregation
   - For each expert $E_j \in E$ and each alternative $A_p \in A$ calculate
   $$F_B(A_p, E_j) = \sum_{q=1}^{n} \mathrm{Card}(N_{pq}^j) \tag{1}$$
   - For each expert $E_j \in E$ order the alternatives $A_1, \dots, A_n$ according to the following rule
   $$A_p P_j A_q \quad \text{iff} \quad F_B(A_p, E_j) > F_B(A_q, E_j) \tag{2}$$
   $$A_p I_j A_q \quad \text{iff} \quad F_B(A_p, E_j) = F_B(A_q, E_j) \tag{3}$$
2. Second-level aggregation
   - For all alternatives $A_p \in A$ calculate
   $$f_B(A_p) = \sum_{q=1}^{n} \mathrm{Card}(M_{pq}) \tag{4}$$
   - Order all the alternatives according to the rule
   $$A_p P A_q \quad \text{iff} \quad f_B(A_p) > f_B(A_q) \tag{5}$$
   $$A_p I A_q \quad \text{iff} \quad f_B(A_p) = f_B(A_q) \tag{6}$$

And the process of aggregation according to Approach II, again with help of modified Borda function, will be described by following steps:

1. First-level aggregation
   - For each criterion $C_i \in C$ and each alternative $A_p \in A$ calculate
   $$F_B(A_p, C_i) = \sum_{q=1}^{n} \mathrm{Card}(M_{pq}^i) \tag{7}$$
   - For each criterion $C_i \in C$ order the alternatives $A_1, \dots, A_n$ according to the following rule
   $$A_p P^i A_q \quad \text{iff} \quad F_B(A_p, C_i) > F_B(A_q, C_i) \tag{8}$$
   $$A_p I^i A_q \quad \text{iff} \quad F_B(A_p, C_i) = F_B(A_q, C_i) \tag{9}$$
2. Second-level aggregation

- For all alternatives $A_p \in A$ calculate

$$f_B(A_p) = \sum_{q=1}^{n} \mathrm{Card}(N_{pq}) \tag{10}$$

- Order all the alternatives according to the rule

$$A_p P A_q \quad \text{iff} \quad f_B(A_p) > f_B(A_q) \tag{11}$$

$$A_p I A_q \quad \text{iff} \quad f_B(A_p) = f_B(A_q) \tag{12}$$

Remark 1. Copeland function is modified analogically. For example, the first-level aggregation according to Approach I by Copeland function is calculated as $F_{Cop}(A_p, E_j)$, where

$$F_{Cop}(A_p, E_j) = \sum_{q=1}^{n} \mathrm{sgn}(\mathrm{Card}(N_{pq}^j) - \mathrm{Card}(N_{qp}^j)), \tag{13}$$

and the second-level aggregation is calculated as $f_{Cop}(A_p)$, where

$$f_{Cop}(A_p) = \sum_{q=1}^{n} \mathrm{sgn}(\mathrm{Card}(M_{pq}) - \mathrm{Card}(M_{qp})). \tag{14}$$

3 Application of two-level aggregation

The behavior of two-level aggregation approaches (Approach I and Approach II) will be examined on the examples.

Example 2. Let us imagine a competition. Let $A_1, \ldots, A_4$ be contestants, which are supposed to be ordered with respect to their performances in challenges C_1, C_2, C_3. Furthermore, let E_1, E_2, E_3 be experts, which are willing to help us. After each challenge, all the experts evaluate all the contestants by points 1, 2, 3, 4 or 5 (5 points should correspond to a really good performance). The opinions of the experts are given in Table 1.

	C_1			C_2			C_3		
	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3
A_1	3	3	3	4	3	1	1	1	3
A_2	1	3	3	2	4	1	1	4	3
A_3	5	3	1	2	1	4	1	2	4
A_4	3	4	1	1	3	4	4	2	1

Table 1 Four contestants $A_1, \ldots, A_4$ evaluated by three experts E_1, E_2, E_3 with respect to their performance in three challenges C_1, C_2, C_3. (Example 2)

If we sum up all the points given to each contestant, we obtain numbers 22, 22, 23 and 23. Who is the winner? If the goal is to rank the alternatives, the absolute values of the points should not be so important. Based on given points we can deduce 9 particular orderings $\mathcal{O}_{ji}$, $i = 1, 2, 3$, $j = 1, 2, 3$. Therefore, we applied the two-level aggregation method mentioned above. The aggregations were calculated with help of modified Borda and Copeland functions and their combinations (Borda function applied to the first-level aggregation while Copeland function applied to the second-level and vice versa). The results are given in Table 2.

At first let us focus on the first column of Table 2. Here two different orderings can be seen despite of the fact, that in both levels of aggregation the same function was applied (Copeland). Although the only difference is in the order of the aggregations (Approach I vs Approach II), the results differs completely. While according to Approach I the winner is A_3, Approach II placed the contestant A_3 in last place in behalf of contestant A_2. The results displayed in the second column of the table are similarly disappointing. Although both approaches employed Borda function for aggregation, the final orderings differ quite significantly. The first conclusion can be stated: The final ordering of the contestants depends on the order of the aggregations.

Approach I	F_{Cop}, f_{Cop}	F_B, f_B	F_B, f_{Cop}	F_{Cop}, f_B
final ordering	$A_3PA_1IA_2PA_4$	$A_2IA_3PA_4PA_1$	$A_3PA_2PA_1IA_4$	$A_3PA_2PA_1IA_4$
Approach II	F_{Cop}, f_{Cop}	F_B, f_B	F_B, f_{Cop}	F_{Cop}, f_B
final ordering	$A_2PA_1PA_4PA_3$	$A_4PA_1IA_2PA_3$	$A_4PA_2PA_1IA_3$	$A_2PA_1PA_3IA_4$

Table 2 Final orderings of the contestants obtained by two approaches and two social choice functions.(Example 2)

The last two columns display the results of combined aggregation, where after the first-level aggregation computed by Copeland function, the second-level aggregation was calculated with help of Borda function and vice versa. Here the results obtained by Approach I are identical and very similar to those in the first two columns. All of them agree on the contestant A_3 as a winner and put the contestants A_1 and A_4 in last places. Aggregating according to Approach II we have obtained two different orderings. Those two different results correspond to two different functions applied in the first level of aggregation. If it was Copeland function, which was employed in the first-level aggregation, the winner seemed to be the contestant A_2. If the first-level aggregation was computed by Borda function, the contestant who was denoted as the best one was the contestant A_4. Does the first-level aggregation method influence the result so much?

The strange behavior observed in Example 2 inspired us to create Example 3, which could explain the process of two-level aggregation a bit more.

Example 3. Let us consider a minor simplification of Example 2. Here the experts are supposed to choose a winner out of only two contestants. The whole competition is described in Table 3.

	C_1			C_2			C_3		
	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3
A_1	5	5	5	1	1	1	1	1	1
A_2	5	1	1	5	1	1	5	1	1

Table 3 Two contestants A_1, A_2 evaluated by three experts E_1, E_2, E_3 with respect to their performance in three challenges C_1, C_2, C_3. (Example 3)

This time the results of two-level aggregation are not influenced by the type of social choice function, all of them provide us the same orderings. Nevertheless, the final ordering depends on the order of aggregations. According to Approach I the winer is the contestant A_1, whereas Approach II points to the contestant A_2 as a winner. Although both the contestants ended the competition with the same scores (21 points), their performances were not the same. The result given by Approach I could be explained like this: According to the expert E_1 the contestant A_2 was better than the contestant A_1, because she won two challenges out of three. The remaining two experts shared the opposite opinion: both of the contestants were on the same level except the first challenge, where the contestant A_1 performed better, i.e. the winner should be the contestant A_1. The expert E_1 was outvoted and the winner is A_1.

We can explain the result of Approach II as well: Obviously, the contestant A_1 won the first challenge, because two experts agreed on that and the third one considered both the contestants as equal. In remaining two challenges one of the experts always preferred the contestant A_2 to the contestant A_1 and the rest of the experts evaluated the performances of both contestants equally and could not decide which one should win. Therefore, the winner of the second and third challenge was the contestant A_2. As the contestant A_2 won two out of three disciplines, it made her a winner of whole competition.

Example 4. Let us have a look at other competition. Again, there are four contestants $A_1, \ldots, A_4$ evaluated by three experts E_1, E_2, E_3 with respect to their performance in three challenges C_1, C_2, C_3. And again, the performances of the contestants are evaluated with help of points (see Table 4).

This time the total scores of the contestants differ. The sums of all points for particular contestants A_1, A_2, A_3 and A_4 are 23, 24, 20 and 21 points. According to these overall scores, the ordering of the contestants is obvious: $A_2PA_1PA_4PA_3$, i.e. the winner is the contestant A_2, whose performance in the first challenge was very good and in the rest of challenges was not so bad. All of the experts agreed on

	C_1			C_2			C_3		
	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3
A_1	2	1	3	3	3	2	2	4	3
A_2	4	4	4	2	2	2	2	2	2
A_3	1	1	1	4	1	3	4	1	4
A_4	3	2	3	2	2	4	1	3	1

Table 4 Four contestants $A_1, \ldots, A_4$ evaluated by three experts E_1, E_2, E_3 with respect to their performance in three challenges C_1, C_2, C_3. (Example 4)

these evaluations. On the contrary, the contestant A_3, which is in last place, totally failed in the first challenge, but the performance in the next two challenges was very good. Though, there is one expert who totally disagrees. What will happen, if we leave out the absolute values of the points and focus only on the orderings of the contestants? The answer is in Table 5.

Approach I	F_{Cop}, f_{Cop}	F_B, f_B	F_B, f_{Cop}	F_{Cop}, f_B
final ordering	$A_3PA_1IA_2PA_4$	$A_3PA_1PA_2PA_4$	$A_3PA_1IA_2PA_4$	$A_3PA_1PA_2PA_4$
Approach II	F_{Cop}, f_{Cop}	F_B, f_B	F_B, f_{Cop}	F_{Cop}, f_B
final ordering	$A_1IA_3PA_2IA_4$	$A_1PA_2IA_3PA_4$	$A_1PA_3PA_2PA_4$	$A_1IA_2IA_3PA_4$

Table 5 Final orderings of the contestants obtained by two approaches and two social choice functions.(Example 4)

Surprisingly, according to Approach I, the winner is the contestant A_3, i.e. the contestant who earned the fewest points. The intended winner, the contestant A_2, dropped to third place. If we seek the final ordering with help of Approach II, we find three possible winners A_1, A_2 or A_3. Obviously, the number of acquired points does not need to correspond to a place in the ranking.

4 Conclusion

In the paper, we studied two approaches to solve the multiple criteria group decision making problem. Both of the applied approaches are based on two levels of aggregation and two social choice functions. The principle of each approach was explained and the behavior of these approaches was observed on several examples. On the examples, we demonstrated that the final ordering of the alternatives is dependant on the order of the aggregations (Approach I versus Approach II) and on the function employed to aggregate the particular preferences. The results given by the last example imply that the final ordering of the alternatives computed with help of social choice function could be completely different from the ordering given by the overall scores of the alternatives.

References

[1] Fiala, P.: *Skupinové rozhodování.* (in Czech) VŠE Praha, Praha, 1997.

[2] Figueira, J., Mousseau, V., Roy, B.: ELECTRE methods. In: *Multiple Criteria Decision Analysis: The State of the Art Surveys* (Figueira, J., Greco, S., Ehrgott, M., eds.), Springer Science, Business Media Inc., New York, 2005, 133-162.

[3] Fishburn P. C.: Social choice functions, *SIAM Review*, vol. **16**, no. **1**, (1974), 63-90.

[4] Keeney, R. L., Raifa, H.: *Decisions with multiple objectives: Preferences and tradeoffs.* John Wiley, New York, 1976.

[5] Kilgour, D. M., Eden, C. (Eds.): *Handbook of group decision and negotiation.* Springer Science+Business Media B. V., 2010.

[6] Saaty, T. L.: *The analytic hierarchy process.* McGraw-Hill, New York, 1980.

Generalization of geometric median

Přemysl Bejda

Mathematical-Physical faculty, Charles University, Dept. of Statistics,
Sokolovská 83, Prague 8, Czech Republic
e-mail: premyslbejda@gmail.com

The geometric median is a very robust method for estimating parameters of location. The efficiency of this method is however rather poor. It was shown that the efficiency can be improved by averaging of chosen observations, which lie in some sense around the original estimate. For this purpose we employ the median absolute deviation. The next step to improve the efficiency, which is introduced in the paper, is to employ weights for observations according to their distance from the original estimate and then to compute the weighted mean. This approach is similar to the one of M estimates. Further we deal with a generalization of boxplot for multidimensional case. We suggest procedure based on the mentioned methods. This leads us to looking for quantiles, where we use a similar approach which was established by Koenker and Bassett. Since there is a direct connection between robust statistic and appearance of extreme events in a sense of economic profit we deal with analogy between the value at risk and the observations which were classified as outliers by our method. The simulation study which compares our method with other kinds of estimates is also included.
JEL classification: C14
AMS classification: 91G70

Key words: breakdown point, contamination, geometric median, robustness.

1 Generalized geometric median

The paper considers robust statistical methods which are currently widely used in quantitative methods and operational research. The estimates suggested in the paper can help us to identify economic models. Boxplot (which is in the paper generalized) is employed in almost any analysis of economic data. Further in paragraph 3.3 we suggest a new method for risk management based on our methods. We should also note that robust statistic and risk management are in the narrow connection. This was discussed in [1].

We will proceed with the work from the previous year. All proves from this part can be found in [1]. The current results starts in chapter connecting with M estimates.

We will consider any normed vector space.

The breakdown point was introduced in [3]. The following definition can be found in [6].

Definition 1. Let $\boldsymbol{x}^0 = (x_1, \dots, x_n)$. Consider the following value of an estimator $T_n(\boldsymbol{x}^0)$ of some functional T. From this initial sample $\boldsymbol{x}^0$ we replace m observations by any values, such a new sample we denote $\boldsymbol{x}^m$ and $T_n(\boldsymbol{x}^m)$ the pertaining value of the estimator. The estimator T_n has the breakdown point $\varepsilon_n^*(T_n, \boldsymbol{x}^0) = \frac{m^*(\boldsymbol{x}^0)}{n}$, where $\sup_{\boldsymbol{x}^m} \|T_n(\boldsymbol{x}^m) - T_n(\boldsymbol{x}^0)\| = \infty$. But we will consider as a breakdown point a number (if exists) $\varepsilon^* = \lim_{n\to\infty} \varepsilon_n^*$.

Consider $S_{\boldsymbol{x}_i} = \sum_{j=1}^n \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$.

Definition 2. Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are any observations from some normed vector space. By a geometric median we mean the number

$$\hat{\boldsymbol{x}} = \underset{i=1,\dots,n}{\operatorname{argmin}} \sum_{j=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|. \tag{1}$$

If there are more observations which satisfies (1), we take an average of only two of them, which are different i.e. $\boldsymbol{x}_i \neq \boldsymbol{x}_j$.

Remark 1. By $\|\cdot\|_q$, $q \in \mathbb{N}$ we understand a norm in $\mathbb{R}^p$ such that any $\boldsymbol{x} \in R^p$ satisfies

$$\|\boldsymbol{x}\|_q = \sqrt[q]{|x_1|^q + \cdots + |x_p|^q}.$$

And we denote $\|\boldsymbol{x}\|_\infty = \max_{i=1,\dots,p} |x_i|$.

Proposition 1. *The break down point of geometric median $\hat{\boldsymbol{x}}$ is one half.*

Let us now derive another kind of estimators from the geometric median.

Definition 3. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are any observations from some normed vector space. Let in a set L is one half of observations which distances from geometric median are less than a distance of any other observation which does not lie in the L. By a GM2 estimator we mean the number which is computed as an average of $\lfloor n/2 \rfloor$ observations lying in L. Denote this by $\tilde{\boldsymbol{x}}$.

Proposition 2. *The break down point of $\tilde{\boldsymbol{x}}$ is one half.*

The set L can be employed for constructing different kinds of estimators. We can take for instance once more the median from the set or $r = \max_{\boldsymbol{x}_i, \boldsymbol{x}_j \in L} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ can serve as an estimator of inter quartile range (from this estimator we can, under assumption of normality, easily derive an estimator of variance) etc.

Definition 4. Let $r = \max_{\boldsymbol{x}_i, \boldsymbol{x}_j \in L} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$. $b > 0$. Let the set E contains all observations $\boldsymbol{x}$ such that $\min_{\boldsymbol{x}_i \in L} \|\boldsymbol{x}_i - \boldsymbol{x}\| \leq br = a$.

Proposition 3. *If we make any estimator $\tilde{\boldsymbol{x}}$ from observations in E then the estimator has the break down point one half. It means that the estimator $\tilde{\boldsymbol{x}}$ is a function of the observations from E. The function has to be bounded on any finite interval.*

Remark 2. The last proposition instruct us, how to include more observations and so increase efficiency. But a should not be too large. We can for instance employ theoretical quantiles of a normal distribution to find an appropriate constant a. Let us denote the constant a as a constant of widening.

Remark 3. Thanks to the properties of a norm we can easily show that our estimators (we employ an average on the set L and E respectively) are scale and location equivariant.

1.1 M-estimate

The M-estimates were introduced firstly in [4]. They generalize the maximum likelihood approach. We employ results which are presented in [9]. We describe only the idea, which is useful also for us.

After adding more observations to geometric median and computing the mean from these observations we can proceed in direction of weighting our observations. We will employ iterative approach, which serves well for M-estimates. We are looking for the solution of these equations

$$\sum_{i=1}^{n} W_1\left((\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^\top \hat{\Sigma}^{-1} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})\right)(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}) = \boldsymbol{0}$$

$$\frac{1}{n}\sum_{i=1}^{n} W_2\left((\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^\top \hat{\Sigma}^{-1} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})\right)(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^\top = \hat{\Sigma} \tag{2}$$

where the functions W_1 and W_2 can differ. The $\hat{\Sigma}$ in (2) can be interpreted as a weighted covariance matrix and $\hat{\boldsymbol{\mu}}$ as the weighted mean. With $w_i = W_1\left((\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^\top \hat{\Sigma}^{-1} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})\right)$ we can express $\hat{\boldsymbol{\mu}}$ as a weighted mean with data-dependent weights. The iterative process is as follows. We know $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ from previous step. We compute the weights and employ them for computing new estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$. The crucial question is the choice of the functions W_1 and W_2 and of the initial estimates.

Existence and uniqueness of solutions of (2) were treated in [8] and more generally in [11].

In the next section we describe how to implement this iterative approach in one dimensional case and we employ our estimates as initial.

In [8] was shown that for multidimensional case the approach of M-estimators under some conditions on the functions W_1, W_2 and the distribution function of the observations has break down point always below $\frac{1}{p+1}$, in the case that we do not know the covariance matrix. In this case is better to try another approach like S-estimates see [2].

1.2 Quantiles

Sometimes is important to study quantiles for multidimensional case. For this purpose serves well depth. General discussion about this topic can be found e.g. in [12]. Behind the scope of this article is also a discussion whether our estimate is an example of depth.

We will rather than depth consider the classical approach to the quantiles. I.e. for $\|\cdot\|_1$ we take for each component quantile separately from numbers which appear in the component. For general case of any norm $\|\cdot\|$ we will employ the approach from the article [7]. Let us choose $0 \leq \alpha \leq 1$. Consider further $\boldsymbol{x} \in \mathbb{R}^p$ and put

$$\rho_\alpha(\boldsymbol{x}) = \|(|x_1|\{\alpha \mathsf{I}_{[x_1 \geq 0]} + (1-\alpha)\mathsf{I}_{[x_1<0]}\}, \ldots, |x_p|\{\alpha \mathsf{I}_{[x_p \geq 0]} + (1-\alpha)\mathsf{I}_{[x_p<0]}\})\|.$$

Definition 5. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are any observations from some normed vector space. By an α quantile from the observations we mean the number

$$\hat{\boldsymbol{x}}_\alpha = \underset{\boldsymbol{x} \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\alpha(\boldsymbol{x}_i - \boldsymbol{x}). \tag{3}$$

2 One dimensional case and simulation study

Let us now consider only one dimensional case. We want to compute a mean from observations in E, but we have to gain the set.

Let X is a random variable with a mean μ, a standard deviation σ and let $q_{\mu,\sigma}(\alpha)$ denotes its quantile function. Then for any continuous distribution G (in case that the second moment is finite) the following ratio is constant for every μ and σ

$$K_{G,\alpha} = \frac{q_{\mu,\sigma}(1-\alpha/2) - q_{\mu,\sigma}(\alpha/2)}{q_{\mu,\sigma}(0.75) - q_{\mu,\sigma}(0.25)}$$

This comes from the fact $\sigma q_{0,1}(\alpha) + \mu = q_{\mu,\sigma}(\alpha)$.

Let us further suppose that $X \sim \mathrm{N}(\mu, \sigma)$ and that we generate random sample from X. So we can gain the set L. We construct estimate of $q_{\mu,\sigma}(0.75) - q_{\mu,\sigma}(0.25)$ such that we put $\mathrm{RIQR} = \operatorname{argmax}_L(x) - \operatorname{argmin}_L(x)$. In this manner we construct a robust estimator of inter quartile range RIQR and it has according to the proposition 2 the breakdown point one half. We put the constant of widening as $a = K_{\mathrm{N}(0,1),\alpha}\mathrm{RIQR}$.

It is useful to note that RIQR has very narrow connection to median absolute deviation.

Otherwise than in the definition 4 we take all observations which distance from their median is less than a. Let us denote the set GM3S. The only difference is that we do not take into account all observations from the set L, but only one from these observations. So the proposition 3 stays valid. This approach gives also better results in practise. We can employ other symmetric distributions than normal.

This approach is not appropriate for asymmetrical distributions. In the case of asymmetry we define the bounds for GM3 as follows $a_u = K_{\mathrm{N}(0,1),\alpha}(\operatorname{argmax}_L(x) - \mathrm{MED}(X)) + \mathrm{MED}(X)$ and $a_l = \mathrm{MED}(X) - K_{\mathrm{N}(0,1),\alpha}(\mathrm{MED}(X) - \operatorname{argmin}_L(x))$, where a_u, a_l is upper bound and lower bound respectively for GM3S and $\mathrm{MED}(X)$ is a median from the random sample of the variable X.

Let us now turn to the approach, where we weight the observations to estimate the parameter of location and the variance. As was mentioned in section 1.1 we employ the results from the theory of M-estimators. We employ biweight function as the weight function $W_1(x) = \left[1 - \left(\frac{x}{k}\right)^2\right]^2 \mathsf{I}(|x| \leq k)$. We put $k = 3.44$ it is for prescribed efficiency 0.85 see [9]. Further we put

$$W_2(x) = \begin{cases} \min\left\{1 - (1-x^2)^3, 1\right\}/x^2 & \text{if } x \neq 0 \\ 3 & \text{if } x = 0 \end{cases} \tag{4}$$

We put $\delta = 0.5$. The iterative algorithm will proceed in the steps:

1. Initialize $\hat{\mu}_1 = \mathrm{GM3}$ and $\hat{\sigma}_1 = \frac{\mathrm{RIQR}}{1.3}$. Put $k = 1$.
2. Compute the weights for all observations $i = 1, \ldots, n$: $w_{1k,i} = W_1(\frac{x_i - \hat{\mu}_k}{\hat{\sigma}_k})$ and $w_{2k,i} = W_2(\frac{x_i}{\hat{\sigma}_k})$.
3. Put $\hat{\mu}_{k+1} = \frac{\sum_{i=1}^n w_{1k,i} x_i}{\sum_{i=1}^n w_{1k,i}}$ and $\hat{\sigma}^2_{k+1} = \frac{1}{n\delta}\sum_{i=1}^n w_{2k,i} x_i^2$. Put $k = k+1$ and return to the second step until $k \leq 10$.

We can terminate also if for instance $|\hat{\mu}_{k+1} - \hat{\mu}_k| < \varepsilon \hat{\sigma}_{k+1}$, for some ε. But such a condition could be too time consuming for simulations. We denote the estimate which is given by this algorithm as GMM, but it should be noted that it is just M-estimate. More interesting could be generalization for multidimensional case.

All employed methods and a generation of random samples were implemented in R. For each situation we have 1000 samples and in each sample 100 observations. The estimators used in simulation study are:

mean, median, $\alpha-$winsorised mean, truncated mean, GM2, GM3 and GMM where GM2, GM3 are means from the sets L and GM3S respectively.

During the simulation study we employ these parameters: for the $\alpha-$winsorised mean we truncate 20 percent of observations from each side, for the $\alpha-$truncated mean we truncate 20 percent of observations from each side, for the GM3 $\alpha = 0.01$. parameters for GMM were set in the description of the algorithm.

In the following table is employed a function

$$\sum_{i=1}^{N} |\widehat{\boldsymbol{x}_i} - \mu|,$$

where N is a number of samples (1000), $\widehat{\boldsymbol{x}_i}$ is some estimate (mean, median, etc.) from sample i, and μ is a parameter of location, which is known (it is usually a median of the not contaminated distribution).

In the table is a normal distribution $\mathrm{N}(0, 1)$ contaminated by some other distribution with probability p.

An expression $\mathrm{U}(a, b)$ denotes a uniform distribution on the interval (a, b).

Distribution $p = 5\%$	Mean	Med	Trim	Winsor	GM2	GM3	GMM
$\mathrm{N}(0, 1)$	79	98	84	82	114	84	80
$\mathrm{N}(0, 100)$	194	106	92	124	88	86	84
Cauchy	292	100	87	86	116	86	83
$\mathrm{U}(-10, 10)$	130	105	91	90	122	88	86
$p = 10\%$							
$\mathrm{N}(0, 100)$	261	114	99	98	133	94	93
Cauchy	524	103	90	89	119	88	85
$\mathrm{U}(-10, 10)$	168	111	97	97	127	92	90
$p = 40\%$							
$\mathrm{U}(-20, 10)$	2023	297	580	1121	200	163	181

Table 1: A normal distribution contaminated by a distribution from the first column with a probability p.

3 Illustrations

3.1 Boxplot

In boxplots we can employ instead of IQR the RIQR estimator. As whiskers can be taken borders, which are given by constant of widening and are used for construction of GM3 (the last and the first observation from GM3S). To consider also the asymmetrical distributions we employ the definition of GM3S for asymmetrical distributions. In a case of really bad contamination we are able to classify outliers better.

The figures were presented last year in [1]. The standard boxplot and our method give very similar results in a case of a mild contamination. When the contamination is higher than 25 % then our boxplot is still able to find outliers in comparison to the standard boxplot.

3.2 Multidimensional boxplot

Since our estimate is designed for multidimensional data and the construction of GM2 and GM3 offers a good approach for constructing multidimensional boundaries, we imagine an idea how to generalize boxplot in a fashion of the geometric median into $\mathbb{R}^2$ space. It will not employ boxes on the other hand we generalize an idea of the inter quartile range.

We will proceed in three steps, which will be illustrated by a pictures.

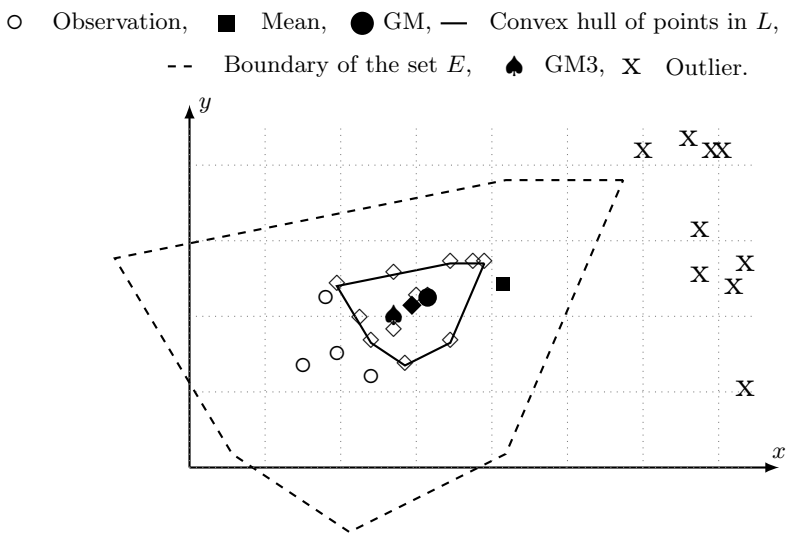
1. In the first step we compute from our data geometric median and for comparison also the mean.

2. We find the half of observations which distance from GM is the lowest according to definition of the set L and GM2. We generate a convex hull of the set L.

3. In the last step we construct the set E. We take the vertices of convex hull from the previous step. Denote this set V. We compute for each vertex $\boldsymbol{v} \in V$ a new point $\boldsymbol{e_v}$, which will lie on the boundary of the set E. We denote the GM estimate as before $\hat{\boldsymbol{x}}$ and a will be the constant of widening which is chosen in the beginning of the procedure and which determine how many observations will be added to E in comparison with L. We compute the $\boldsymbol{e_v}$, such that $\boldsymbol{e_v} = \hat{\boldsymbol{x}} + a(\boldsymbol{v} - \hat{\boldsymbol{x}})$. We found for each $\boldsymbol{v} \in V$ its

$\boldsymbol{e_v}$. We denote this set E_V and find convex hull of this set. This is easy, because we connect the points $\boldsymbol{e_v}$ for which are the points $\boldsymbol{v}$ connected. This new convex hull bounds the set E.

This approach is different from the approach described in the definition 4, but it is easy to show, that we employ just special case. By this approach we are able to grasp the asymmetry in the distribution of the data better. We choose $a = 3.45$.



3.3 Multidimensional value at risk

The last year we have presented the concept of outlying shortfall in [1]. We will continue in this manner. Very brief overview about the topic of VaR and expected shortfall can be found e.g. in [5]. Modern approach to multidimensional value at risk can be found e.g. in [10]. We follow rather different way which should serve as an illustration of our methods in practise.

We utilize the method from paragraph 1.2. Let us assume that we have two series of daily returns $r_{i,t}$ of some stocks S_i, where $i = 1, 2$ and $t = 1, \ldots, T$. We suppose that the returns are independent for each time (they are not necessarily independent in one day between each other). We are interested in daily (for simplicity) value at risk (VaR) at a confidence level 5 %. Let us assume that we have three years of history it means $T = 800$ observations.

The method from paragraph 1.2 is directly applicable only in a case that the time series are independent between each other, otherwise the solution does not take into consideration the dependence structure and so in a case of positive correlation the VaR is too low and in a case of negative correlation too high. We will consider only linear dependence which will be measured by Pearson covariance matrix $\hat{\Sigma}$. On this estimate we apply Cholesky decomposition $\hat{\Sigma} = LL^\top$, in case of $\hat{\Sigma}$ regular. Let $\boldsymbol{r}_t = (r_{1,t}, r_{2,t})^\top$. We compute $\boldsymbol{y}_t^\top = \boldsymbol{r}_t^\top L^{-1}$ and put $Y = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)^\top$ as a $2 \times T$ matrix. After that we apply the formula (3) on the matrix Y. Then we have to multiply resulting vector $\boldsymbol{v}$ by L ($\hat{\boldsymbol{x}}_\alpha^M = L\boldsymbol{v}$) to get the same dependence structure as in the data.

This is equivalnet to the following procedure, where we employ generalized Mahalanobis distance. We can rewrite (3) as

$$\hat{\boldsymbol{x}}_\alpha^M = \underset{\boldsymbol{r} \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_\alpha^M(\boldsymbol{r}_t - \boldsymbol{r}),$$

where

$$\rho_\alpha^M(\boldsymbol{x}) = \|(|y_1|\{\alpha \mathsf{I}_{[y_1 \geq 0]} + (1-\alpha)\mathsf{I}_{[y_1 < 0]}\}, \ldots, |y_p|\{\alpha \mathsf{I}_{[y_p \geq 0]} + (1-\alpha)\mathsf{I}_{[y_p < 0]}\})\| \tag{5}$$

and $\boldsymbol{y}_t^\top = (\boldsymbol{r}_t - \boldsymbol{r})^\top L^{-1}$. In the case of our example $p = 2$.

$r_{1,t}$ has a standard normal distribution $\mathrm{N}(0, 1)$ contaminated by t_4 distribution with a probability 5 %. The series $r_{2,t}$ has a standard normal distribution $\mathrm{N}(0, 1)$ contaminated by t_6 distribution with a probability 10 %. Since we have only two series we consider the correlation coefficient between them as ρ. We will explore different values of ρ. Further we consider the first series with twice higher weight (i.e. we multiply, in the simulated data, the first series by two).

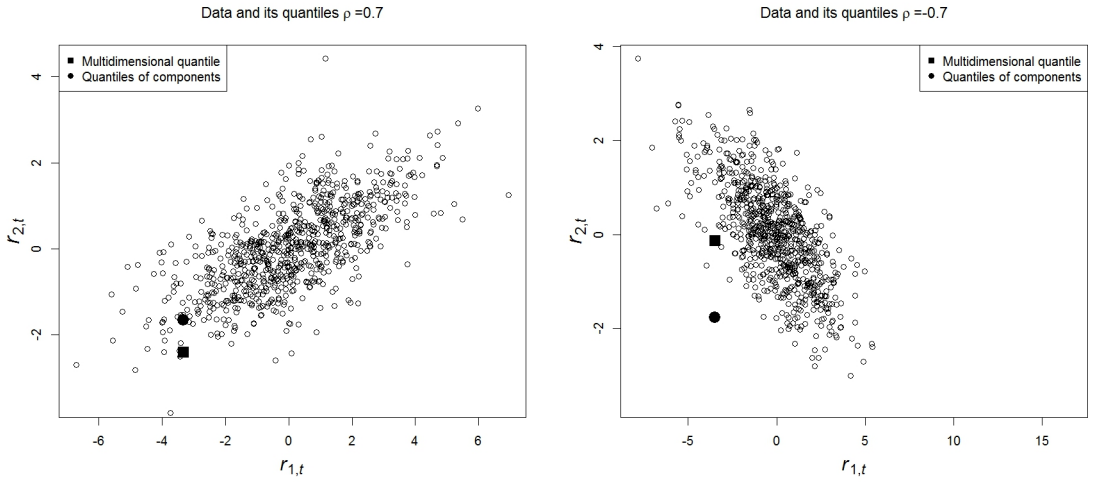


Figure 1: On the pictures are time series simulated with different correlation ρ. We have 800 observations. We estimate 5 % quantile by our method and componentwise.

The first part of the picture demonstrate opposite the situation when correlation is positive. We get for the second component very low value. I.e. when we lost in the first component we should await a loss in the second as well.

The next part of the picture demonstrates an opposite situation. We see that in the case of negative correlation the result is very low for the first component and for the second is almost positive. I.e. according to our method: when a bad scenario takes place and we lost in the first component, the loss in the second component can not be so bad.

References:

[1] Bejda P.: Generalization of geometric median. *In: Mathematical Methods in Economics.* Jihlava, (2013), 25–30.

[2] Davies, P. L.: The asymptotic of Rousseeuw's minimum volume ellipsoid estimator. *The Annals of Statistics* **20.**, 4 (1992), 1828–1843.

[3] Donoho D. L., Huber P. J.: The notion of breakdown point. *In: Bickel PJ, Doksum KA, Hodges JL (eds) A Festschrift for Erich Lehman.* Wadsworth, (1983), 157–184.

[4] Huber P. J.: Robust estimation of location parameter. *The Annals of Mathematical Statistics* **35.**, 1 (1964), 73–101.

[5] Jorion P.: *Financial riskmanager handbook.* JOHN WILEY & SONS, New Jersey, 2003.

[6] Jurečková J., Picek J.: *Robust Statistical Methods with R.* Chapman and Hall/CRC, Boca Raton, 2005.

[7] Koenker R., Bassett G.: Regression quantiles. *Econometrica* **46.**, 1 (1978), 33–50.

[8] Maronna R.: Robust M-Estimators of Multivariate Location and Scatter. *The Annals of Statistics* **4.** (1976), 51–67.

[9] Maronna R., Martin D., Yohai V.: *Robust statistics theory.* JOHN WILEY & SONS, New Jersey, 2006.

[10] Prékopa A.: Multivariate value at risk and related topics. *Annals of Operations Research* **193.**, 1 (2012), 49–69.

[11] Tatsuoka K. S., Tyler D. E.: On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *The Annals of Statistics* **28.**, 4 (2000), 1219–1243.

[12] Zuo Y., Serfling R.: General notions of statistical depth function. *The Annals of Statistics* **28.**, 2 (2000), 461–482.

Detection of Influential Factors on Unemployment Duration of Tomáš Bat'a University Graduates by the Hazard Model

Silvie Bělašková [1], Eva Fišerová[2]

Abstract.

In this retrospective study we analyze the duration of the registration on the Labour Office from the point of view gender problematic and employability of studied faculty conducted by Tomáš Baťa University. We collected data on 512 graduates, from which we obtained the information on age, gender, approbation and the length of stay in the registry on the Labour Office after the graduation, and the statute in the registry (leaving the registry because of work, still in the registry, leaving the registry for another reason, e.g. the own request). The study indicates that the length of stay in the registry affects significantly only gender. The factors age and studied faculty were not statistically significant. Results were obtained by means of survival analysis and the Cox proportional hazards model with right-censored tied data.

Keywords: Unemployment duration, Cox regression model, Survival analysis

JEL classification: J64, R23
AMS classification: 62N01, 62N02, 62P20

1 Introduction

The current high rate of graduates unemployment is a reflection caused by unfavorable economic situation and also the fact that the graduates as a risk group of economically active population are disadvantaged on the labor market due to lack of practice and other factors. The most important factor of employability is the ability of graduates get their first job and keep a job. Therefore, the unemployment rate of graduates is considered as a good measure of their employability in the labor market. However, it is necessary to take into account that employment and unemployment depends not only on the quality of vocational education, see [13]. In this paper, we evaluate the influence of some of these factors as age, gender and studied faculty of Tomáš Baťa University in Zlín in Czech Republic on unemployment duration are. The study aims in particular to establish whether the studied faculty significantly affects the unemployment duration after the graduation.

The Cox proportional regression model [6] is one of the widely used methods of survival analysis for exploring the relationship between the hazard function at the given time (a risk of employment at time t) as a dependent variable, and several explanatory variables, in particular, age as a continuous covariate, gender and studied faculty as categorical covariates. This approach is focused directly on the hazard function which represents the failure time (the unemployment duration) of an individual in the population. The Cox proportional model assumes a continuous hazard without ties, i.e., events cannot occur at the same time. Four modifications of the likelihood

[1]Department of Geoinformatics, Faculty of Science, Palacký University, Olomouc, Czech Republic, belaskova@fai.utb.cz
[2]Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University, Olomouc, Czech Republic, eva.fiserova@upol.cz

function to adjust ties were proposed, namely the discrete method, the Peto–Breslow method, the Efron method and the exact method ([2, 8]).

The survival analysis, see e.g. [12], involves modeling of the time until the occurrence of an event of interest, prototypical such an event is a death, from which the name "survival" analysis comes, although the ambit of applications is much broader. More precisely, in the unemployment duration analysis, getting first job can be considered as an event, then the hazard rate of employability can be determined, and, finally, it is possible to evaluate whether the factors as studied faculty, gender and age are, affect this hazard rate.

In this retrospective study we analyze the unemployment duration from the point of view gender issue and employability of studied approbation conducted by Tomáš Baťa University. We focused on students who successfully graduated at Tomáš Baťa University and after that were registered on the Labour Office. The dataset comes from the Labour Office of Zlín region which collect the information about unemployed people. The date of graduates registration represents the beginning of our time period (time to event), the time of their leaving from the registry is considered as the end of this period. In the survival analysis, it is a common situation, called as right censoring, that for some subjects it is known only the date of registration. To complete the information about employment of subjects, the indicator censored variable is used. The symbol 1 denotes a new employer (a subject gets a new job), 0 means that a subject is still in the registry or had left the registry for another reason, e.g. at his own request. As we already mentioned, as covariates we have considered age as a continuous variable, and gender and graduated faculty as categorical ones.

2 Survival analysis

Let T be a random variable denoting time to event, i.e., time from registration on the Labour Office to the first getting job in our study. The length of time that graduates remain unemployed after a registration can be measured by the survival function. The survival function S is the probability that time of the event is later than specified time t, i.e., $S(t) = P(T > t)$, $t \in \langle 0, \infty)$. The Kaplan and Meier estimator [10], also called as the product limit, is a well known estimator for estimating the survival function. This method is suitable even for censored data, particularly right-censoring, which occurs if a graduate had left the registry from another reason or is still in the registry after the end of the study. Let $t_{(1)} < t_{(2)} < \cdots < t_{(m)}$ denote the distinct ordered times of getting the first job (without censoring times). Let d_i is the number of getting the first jobs at $t_{(i)}$, and n_i is the number of unemployed before $t_{(i)}$. Then the Kaplan-Meier estimator of the survival function is

$$\widehat{S}(t) = \prod_{i:t_{(i)}<t} \left(1 - \frac{d_i}{n_i}\right) \tag{1}$$

The Kaplan-Meier estimate is a step function with the points of discontinuity (jumps) at the observed getting job times.

The hazard function is defined as the event rate at time t conditional on the survival until time t or later, i.e.,

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t}, \tag{2}$$

what can be rewritten as

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}, \tag{3}$$

where $f(t)$ denotes the density function of the lifetime distribution and $F(t)$ means the lifetime distribution function. The hazard function is non-negative and represents a failure time of an

individual in the population. The Cox proportional regression model specifies the hazard rate at "survival" time t for an individual with covariates $\boldsymbol{x} = (x_1, \ldots, x_k)'$ in the form

$$h(t, \boldsymbol{x}) = h_0(t) \exp(\beta_1 x_1 + \cdots + \beta_k x_k), \tag{4}$$

where $h_0(t)$ is the baseline hazard function when all covariates are zero, x_i is the ith covariate in the model, and β_i is the regression coefficient for the ith covariate, x_i.

Cox (1972) suggested estimation of regression coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ based on the "partial" likelihood function

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{\beta}' \boldsymbol{x}_i)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}' \boldsymbol{x}_j)} \right)^{\delta_i}, \tag{5}$$

where $R(t_i) = \{j : t_j \geq t_i\}$ denotes the risk set at the time t_i. The risk set $R(t_i)$ of individuals whose event times exceed t_i is a convenient mechanism for excluding from denominator those individuals who already experienced the event and from this point of view are not part of this risk set (Allison, 1995). The symbol δ_i is an indicator for censoring, where 1 denotes censored and 0 event. The function L is called a "partial" likelihood function because it considers only probabilities for failed subjects. When times in the continuous time model are grouped, ties in failure times can be observed. In the Cox partial likelihood model ties are not considered, because the formula (5) is valid only for data which are not grouped. If the number of observation and of ties is tolerable with respect to computing time, Breslow [3] proposed the following approximation for estimating regression coefficients. Suppose the events occur at N distinct times $t_1 < t_2 < \cdots < t_N$. Let us denote d_i the total number of failures at the time t_i, D_i set of all subjects who fail at the time t_i and $\boldsymbol{s}_i$ sum of covariate values over all subjects in the set D_i, that is $\boldsymbol{s}_i = \sum_{j \in D_i} \boldsymbol{x}_j$. The Breslow partial likelihood function is given as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \frac{\exp(\boldsymbol{\beta}' \boldsymbol{s}_i)}{\left(\sum_{j \in R(t_j)} \exp(\boldsymbol{\beta}' \boldsymbol{x}_j) \right)^{d_i}}. \tag{6}$$

3 Unemployment study

Let us consider the problem, whether chosen covariates affect time to event (a length of stay in the registry on the Labour Office) of student after the graduation. In this retrospective study, 512 successfully graduates of faculties of Tomáš Baťa were reviewed in the period from January 1, 2011 to March 25, 2014. From the Faculty of applied informatics (FAI), there were 66 students (42 have left the registry and 24 are still in the registry or have left from another reason); from the Faculty of management and economics (FAME), 255 students were observed and from this count 174 have left the registry and 81 are still registered; from the Faculty of humanities (FHS), 65 students were registered and 49 have left the registry and 16 are still in the registry; from the Faculty of technology (FT), 80 students were registered, 59 have left the registry and 21 are still in the registry; the Faculty of multimedia communication (FMK) where from 40 students registered after graduation 28 left the registry and 12 are still in the registry, and the last one is the University institute (UI) where 6 students were registered and 2 have left the registry yet. The statistical software package SAS Version 9.3 (SAS Institute, NC, USA) and R version 2.15.1 were used for all statistical analysis.

Time to event is considered as the response variable, the event is getting a job. This "survival" time is considered as number of days in the registry. In general, we observed $N = 512$ graduates, 341 female (26.98 percent censored) and 171 male (38.60 percent censored). The female graduates has average number of days in the registry $m_f = 207.48$ with the standard error $s_f = 12.57$; for the male graduates $m_m = 298.76$ and $s_m = 27.65$ days. The Kaplan-Meier estimates of survival functions for male and female with 95% confidence interval are demonstrated in the Figure 1. The confidence interval are computed using the Hall-Wellner approach which is a natural extension

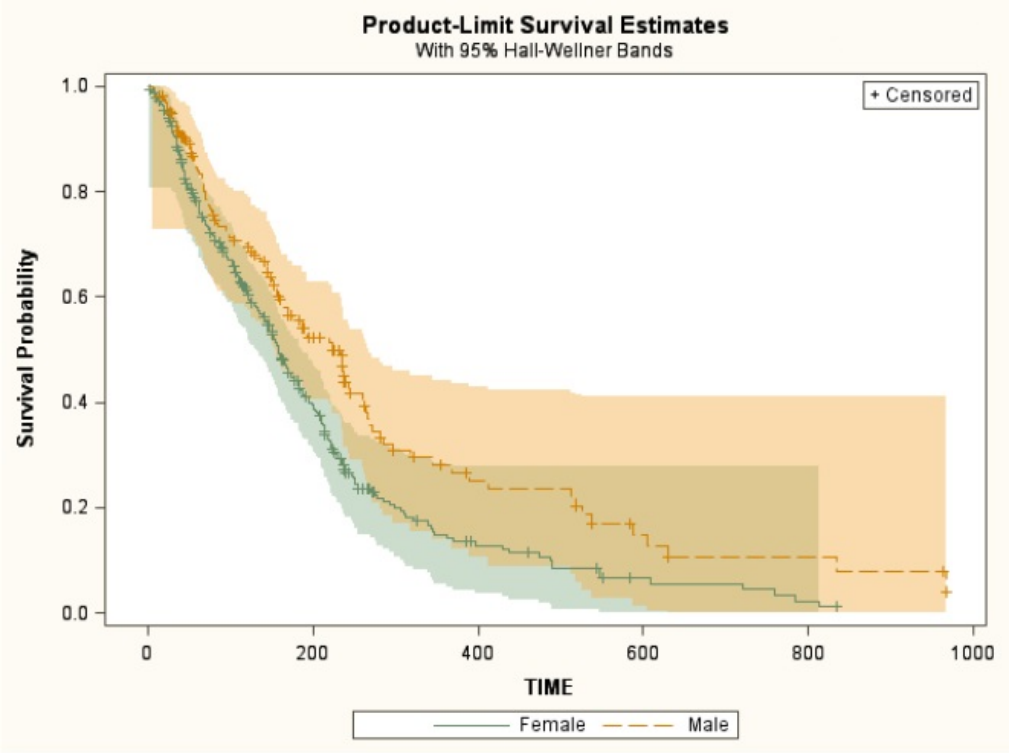


Figure 1 The Kaplan-Meier survival curves grouped by the variable gender.

of Kolmogorov-Smirnov simultaneous confidence band for survival curve using the Kaplan-Meier estimator, for more detail see [9]. These step functions imply that female has the worse survival experience, in our case it means that a length of time spend on the Labour Office is shorter for female than for male. We can see that 50 percent of female gets a job after 160 days and 75 percent of female gets their first job after 250 days while males get their first job in 50 percent after 230 days and after 420 days in 75 percent.

Now we will focus on the question whether the unemployment duration depends also on other factors as age and studied faculty are. The graduates from the Faculty of management and economics (FAME) has the largest average number of days in the registry, namely $m_{\max} = 249.19$ days with the standard error $s_{\max} = 18.14$ days. On the contrary, the least average number of days in the registry have graduates from the University Institute (UI), $m_{\min} = 78.67$ and $s_{\min} = 13.34$ days. For other faculties, the results are following: the Faculty of applied informatics (FAI) $m = 229$, $s = 26.24$; Faculty of humanities (FHS) $m = 184.99$, $s = 22.18$; Faculty of multimedia communication (FMK) $m = 203.14$, $s = 32.4$; and Faculty of technology (FT) $m = 218.08$, $s = 24.8$. The age of graduates across of the studied faculty is almost similar. The average age is $m = 23.94$ with standard deviation $s = 1.26$. The results of the Cox proportional regression model are presented in the Table 1. One can see that the length of the stay in the registry of the Labour Office significantly affects only gender (p-value 0.0006). The significance of covariates is verified by the Wald chi-square statistics with one degrees of freedom. Further, the results indicate that females have 59% (hazard ratio 1.589) increase in the hazard rate compared to males. It means that females have about 59% greater chance that they leave the registry earlier than males. The statistical significance of used additive model without interactions was tested by the likelihood ratio test (p-value is 0.0188), the score test (p-value is 0.0215) and the Wald test (p-value is 0.0228). Proportionality of covariates were tested by interaction with time covariate and graphically checked by the survival function. A model with interactions was also analyzed but this model was not statistically significant.

Coefficient	Estimate	Stand. error	p-value	Hazard Ratio
Age	-0.01001	0.04504	0.8241	0.990
FAI	-0.01397	0.73111	0.9848	0.986
FAME	-0.28550	0.71565	0.6899	0.752
FHS	-0.04361	0.72469	0.9520	0.957
FMK	0.00716	0.73636	0.9922	1.007
FT	-0.09209	0.72501	0.8989	0.912
Female	0.46310	0.13564	0.0006	1.589

Table 1 The results of estimation in the Cox proportional regression model.

4 Discussion

One of the most discussed phenomenon of the employability of the graduates is whether the choosing of studied faculty affects the employability of these graduates. We evaluated the affect of gender, studied faculty and age of graduates of Tomáš Baťa University, which are available from the database of the Labor office, on the hazard function of the length of stay in the registry of the Labour Office. Our results indicate that the hazard function is not affected by age and studied faculty but is affected by gender. In particular, females have 59% (hazard ratio 1.589) increase in the hazard rate compared to males and this increase is statistically significant. It means that females have 59% greater chance that they leave the registry earlier than males.

Acknowledgements

The authors thankfully acknowledge the support by the Operational Program Education for Competitiveness - European Social Fund (project CZ.1.07/2.3.00/20.0170 of the Ministry of Education, Youth and Sports of the Czech Republic) and the grant IGA_PrF_2014028 of the Internal Grant Agency of Palacky University Olomouc. The authors are grateful to the Ing. Ivona Macurková from the Labour Office for helpful comments and collecting the data.

References

[1] Allison, P. D. and SAS Institute: *Survival Analysis Using the Sas System: A Practical Guide.* SAS Institute Incorporated,1995. http://books.google.cz/books?id=xaBRKg5mtoIC.

[2] Breslow, N.E.: *Discussion of Professor Cox's Paper*, J. Royal Stat.Soc. B, 34,1972. 216/217.

[3] Breslow, N. E.: *Covariance Analysis of Censored Survival Data*, Biometrics, 30,1974, 89–99.

[4] Cary, N.C. SAS Institute Inc.: *User's Guide*, SAS Institute Inc. SAS/STAT® 9.2, 2008.

[5] Collett, D.: *Modeling Survival Data in Medical Research*, London: Chapman & Hall, 1994.

[6] Cox, D.R.: *Regression Models and Life Tables*, Journal of the Royal Statistical Society, Series B, 20, 1972, 187–220, with discussion.

[7] Cox, D.R. and Oakes, D.: *Analysis of Survival Data*, London: Chapman & Hall, 1984.

[8] Efron, B.: *The efficiency of Cox's likelihood function for censored data*, Journal of American Statistical Association, 72, 1977, 557– 565

[9] Hall, W.J. and Wellner, J.A.: *Confidence Bands for a Survival Curve from Censored Data*, Biometrica, 67, No.1, 1980, 133-143

[10] Kaplan, E.L. and Meier, P.:*Nonparametric Estimation from incomplete observations*, Journal of the American Statistical Association, 53, 1958, 457-481

[11] Klein, J.P. and Moeschberger, M.L.: *Survival Analysis: Techniques for Censored and Truncated Data*, Springer, New York, 1997.

[12] Kleinbaum, D.G.and Klien, M.: *Survival Analysis: A Self-Learning Text. Second Edition*, Springer-Verlag, New York, 2005.

[13] Koucký, J. and Zelenka, M.: *Postavení vysokoškoláků a uplatnění absolventů vysokých škol na pracovním trhu*, 2011.

A Belief Theoretic Approach to Finding a True Value from Recommendations in E-business

Ladislav Beránek[1]

Abstract. Recommendation systems have become a common way how to help people when they have to decide in complex selections. Recommender systems have become an integral part of many e-business applications. Various algorithms have been proposed for recommendations and many business solutions are built for different applications. The algorithms are usually based on models of user preferences, preference relations and their learning. Nevertheless, the best approach does not exist here due to the high complexity and uncertainty of the problem. Therefore, a recommendation system typically includes modules with multiple recommender algorithms and the resultant recommendation is created by combination of more partial results. In this paper, we deal with the determination of the grand truth within solving recommendation problem. We use the framework of Dempster-Shafer theory with focus on information fusion. The experiments results show that our proposed method leads to improvements of recommendation over the baseline models.

Keywords: Recommendation, e-business, Dempster-Shafer theory, user preferences, information fusion, ground truth.

JEL Classification: C88
AMS Classification: 91B42

1 Introduction

A large number of users share and process all sorts of data within various applications based on the Internet infrastructure. Users evaluate various products and express their opinions to different events. Tripadviser.com server might be an example. Users can evaluate here certain hotel on the base of their satisfaction with its services. Wikipedia provides feedback tool to engage the reader in a review of article quality based on four criteria, i.e., ”trustworthy”, ”objective”, ”complete” and ”well written”. Such activity is referred also as users’ recommendation. A large number of users perform their evaluation or classification of certain product or services. As this method is useful, organizers usually have little control over quality of users’ activity. Reaction of individual users may vary substantially, and in some cases, they may even be controversial. The question is then how to integrate feedback from multiple users to get an objective opinion. Commonly used heuristics such as ”majority voting” and ”take the average” ignore individual user experience and can fail, for example in an environment where there are users with malicious intent, who want even to harm certain product or service. The aim of this paper is to propose and to test a method to determine the grand truth for recommendation without knowing the previous experience of users. For this purpose it is used an approach based on the Dempster-Shafer theory.

Within this theory, the operation discounting is defined. At this operation, the value of belief function varies in dependence on certain additional information concerning to recommendation or if the pieces of information, to be integrated, are contradictory. When it is necessary to decides to implement discounting process the following questions are to be solved: What resources (recommendation) are to be discounted? Up to what extent these resources should be discounted? The model used in this paper introduces an iterative method which automatically determines the discount rate on the base of the reliability of sources. The advantage of this approach is that it does not require any additional meta-information about the reliability of sources of recommendation. The method assumes only that the more specific source of information is in conflict with the majority opinion, the stronger this source must be discounted.

The rest of this paper is organized as follows. Section 2 provides an overview of related work. Section 3 formulates the problem and introduces a belief function framework with the proposed model. Section 4 presents experimental results on synthetic data. Conclusions are composed in Section 5.

[1] Faculty of Economics, University of South Bohemia, Studentska 13, Ceske Budejovice, Czech Republic
e-mail: beranek@ef.jcu.cz.

2 Related Works and Background

Currently, a number of studies deal with the setting involving multiple recommenders. For example, the work such as [4], [5], [7], [12], [16], [19], [20], [26], [27], [29] focus on the estimating the error rates of recommenders. Authors [5] deal with selecting the best set of all available information from users for model training. These works focus on learning classifiers directly from user data instead of estimating true value. Papers [18], [20] use a probabilistic framework for solving classification, regression and ordinal regression problem with multiple annotators. This framework is based on the assumption that the expertise of each recommender does not depend on these data. Works [29],[31], [32] develop this approach, but do not build fully on this premise. There are some other related work, which focuses on a different setting [4], [30]. Recent work [11] pays attention to regression problem under multiple recommenders or observers, with the use of less parametric methods for modeling and designing observers regression function. The use of belief functions to the aggregation of multiple sources (recommenders) is described in [9]. However the author's approach is quite general and they do not specify any application. The paper [15] is one of the first analyzes of conflict information in the theory of belief functions. Authors in the paper [14] modify Dempster rule for aggregating of conflicting evidence. They introduce a measure of distance of conflicting evidences. This measure is then used in modification of Dempster rule. The similar approach is used in the work [10]. Authors describe the problem of aggregation of conflicting evidence in the theory of belief functions and introduce another measure of conflict between evidences. Interesting work is [3], which describes the influence of conflicting information in the field of robotics and proposes metrics to solve the described problems. Similar problem in sensor networks is analyzed in [21]. Schubert [23] introduced an iterative approach for calculating discount rates. He calculates degree falsity as the degree of conflict between evidences (information sources). Then each piece of evidence is discounted in proportion to the degree that it contributes to the conflict. Discounting is performed in a sequence of incremental steps, with conflict updated at each step, until the overall conflict is brought down to a predefined acceptable level.

2.1 Belief Function Theory

Information related to decision making about trust is often uncertain and incomplete. Therefore, it is of vital importance to find a feasible way to make decisions under this uncertainty. The desirable properties of trust representations in an Internet auction (and in all online systems) are:

- Trust representation should integrate different types of uncertainty:
  - Uncertainty about the outcome of a transaction and uncertainty resulting from the fact of using a second-hand experiences.
- Trust representation should allow for decision making and should have the following properties:
  - Possibility to rank alternatives;
  - Possibility to compare with own standards.

Our model is a particular application of the Dempster-Shafer theory. The Dempster-Shafer theory [22] is designed to deal with the uncertainty and incompleteness of available information. It is a powerful tool for combining evidence and changing prior knowledge in the presence of new evidence. The Dempster-Shafer theory can be considered as a generalization of the Bayesian theory of subjective probability. In this paper, we propose a unique trust model based on the Dempster-Shafer theory which combines evidence concerning reputation with evidence concerning possible illegal behavior on an Internet auction.

In the following paragraphs, we give a brief introduction to the basic notions of the Dempster-Shafer theory (frequently called theory of belief functions or theory of evidence).

Considering a finite set referred to as *the frame of discernment* Ω, a *basic belief assignment (BBA)* is a function m: $2^{\Omega} \rightarrow [0,1]$ so that

$$\sum_{A \subseteq \Omega} m(A) = 1, \tag{1}$$

where $m(\varnothing) = 0$, see [22]. The subsets of 2^{Ω} which are associated with non-zero values of m are known as *focal elements* and the union of the focal elements is called *the core*. The value of $m(A)$ expresses the proportion of all relevant and available evidence that supports the claim that a particular element of Ω belongs to the set A but not to a particular subset of A. This value pertains only to the set A and makes no additional claims about any subsets of A. We denote this value also as a *degree of belief* (or *basic belief mass – BBM*).

Shafer further defined the concepts of *belief* and *plausibility* [22] as two measures over the subsets of Ω as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B), \tag{2}$$

$$Pl(A) = \sum_{B \cap A \neq \phi} m(B). \tag{3}$$

A *bba* can also be viewed as determining a set of probability distributions P over Ω so that $Bel(A) \leq P(A) \leq Pl(A)$. It can be easily seen that these two measures are related to each other as $Pl(A) = 1 - Bel(\neg A)$. *Moreover both of them are equivalent to m.* Thus one needs to know only one of the three functions *m*, *Bel*, or *Pl* to derive the other two. Hence we can speak about belief function using corresponding *bba*s in fact.

Dempster's rule of combination can be used for pooling evidence represented by two belief functions Bel_1 and Bel_2 over the same frame of discernment coming from independent sources of information. The Dempster's rule of combination for combining two belief functions Bel_1 and Bel_2 defined by (equivalent to) *bba*s m_1 and m_2 is defined as follows (the symbol $\oplus$ is used to denote this operation):

$$m_1 \oplus m_2(A) = \frac{1}{1-k} \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \ , \tag{4}$$

where

$$k = \sum_{B \cap C = \varnothing} m_1(B) \cdot m_2(C) \, . \tag{5}$$

Here k is frequently considered to be a *conflict measure* between two belief functions m_1 and m_2 or a measure of conflict between m_1 and m_2 [22]. Unfortunately this interpretation of k is not correct, as it includes also internal conflict of individual belief functions m_1 and m_2 [11], [12]. Dempster's rule is not defined when $k = 1$, i.e. when cores of m_1 and m_2 are disjoint. This rule is commutative and associative; as the rule serves for the cumulation of beliefs, it is not idempotent.

When calculating contextual discounting we also use the un-normalized (conjunctive) combination rule established by Smets [25] in this form (we use the symbol $\otimes$ to denote this operation):

$$m_1 \oplus m_2(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \, . \tag{6}$$

Belief Function Correction

When receiving a piece of information represented by a belief function, some metaknowledge regarding the quality or reliability of the source that provides information, can be available. In the following paragraphs, we describe briefly some possibilities how to correct the information according to this metaknowledge.

To handle the lower reliability of information sources, a discounting scheme has been introduced by Shafer [22]. It is expressed by equations:

$$^{\alpha} m(A) = \begin{cases} (1-\alpha) \times m(A) & \text{if } A \subset \Omega \\ \alpha + (1-\alpha) \times m(\Omega) & \text{if } A = \Omega \end{cases} \tag{12}$$

where $\alpha \in [0,1]$ is a discounting factor and $^{\alpha} m(A)$ denotes the discounted mass of $m(A)$. The larger α is, the more masses are discounted from $A \subset \Omega$, while the more mass is assigned to the frame of discernment Ω.

An extension of this classical approach is contextual discounting. It is described in detail in referenced papers [6], [7]. Contextual discounting is used if we know that the reliability of predictions for the assignment of individual elements of the framework differs.

3 The Suggestion of Our Model

Traditional data fusion processing based on Dempster/Shafer theory consists of obtaining of BBAs due to some mathematical model in the first step. The second step is the discounting of some BBAs which we know about that they are less reliable (6). The final step is the integration of BBAs using a Dempster rule (4) or using some other suitable combination rule [23], [24], [25], [28]. As it was described above discounting process is used when we have meta-information about the reliability of some contextual sources of information (BBA) and it is

necessary to have some approach how to express the value of discounting factor [1], [2]. In the most cases, the discount rate is adjusted manually, but some authors have suggested several methods how to obtain them automatically. In [25], Smets calculates the discount factor by minimizing the error function. This method focuses on the classification of data and requires a set of labeled data. Martin et al. establish in [17] the discounting factor evaluation method that is based only on the values of BBA themselves. Similar approach which is the basis of our work is presented in [13].Defining what the majority opinion means within the Dempster-Shafer theory is not easy. Murphy [18] for example suggested using average BBAs and argued that the average properties are better suited for the fusion of contradictory evidence:

$$m_{mean} = \frac{1}{M}\sum_{i=1}^{M} m_i \tag{7}$$

This opinion is valid considering the fact that if subset s1 from S corresponds to the cluster of concordant BBAs and if this subset contains more BBAs than any other cluster, then m_{mean} will probably be closer to BBAs forming the s1. Hence m_{mean} can be used as an estimate of the majority opinion [17]. We therefore propose to review the first set of discount factors by the following way:

$$\alpha_i^0 = d_{BPA}(m_i, m_{mean}) \tag{8}$$

Where d_{BPA} is defined subsequently [17]:

$$d_{BPA}(m_1, m_2) = \sqrt{1/2(\vec{m}_1 - \vec{m}_2)^t D(\vec{m}_1 - \vec{m}_2)} \tag{9}$$

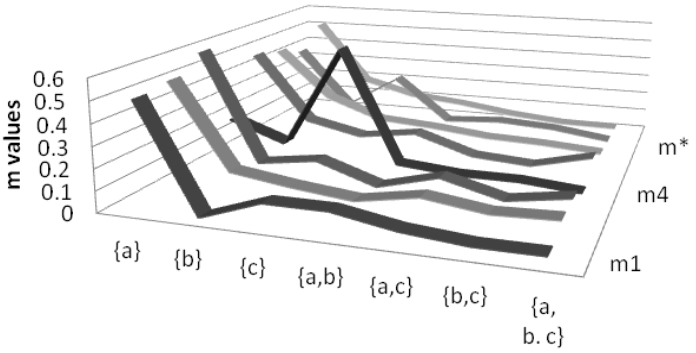
Here, the $\vec{m}$ is *BBA* expressed in the form of vector and D is the matrix which has dimensions $2^N \times 2^N$ with elements $D(A,B)=A\cap B|A\cup B$. Equation (10) gives low values of discounting factor α for *BBA*s near to the mean (they are in accordance with the opinion of the majority) and a high degree of discounting factor for *BBA*s that differ considerable from the mean (the ones that are the cause of disagreement).

In this paper, we use an iterative method for calculating of discounting factors. In the first step, discounting factors are calculated for each member of the initial settings using equation (8). Then this iteration process is applied on the *BBA*s set S_1. New values of discounting factors are obtained. This iteration is repeated and the value of discounting factors increases but more and more slowly. To determine the optimal set of discount factors among those computed at each iteration step a posteriori analysis is employed. We investigate the conjunctive combinations obtained at each step, and compare them with categorical *BBA*s by distance d_{BPA}. Iteration that gives minimum distance is optimal number of iteration i_{opt}. Relative values of discount factors in single steps affect the result of the result of information fusion process as much as the absolute value. In other words, it is not sufficient to have a high degree of value on unreliable sources, it is also necessary that the measure of the difference between reliable and unreliable sources be large enough. Therefore we perform the optimum setting of values α_i using iteration. We calculate a discounting factor of the initial set of *BBA*s and then recalculate new values of *BBA*s of this set. This process is repeated as described in the previous paragraph. Consecutive values of discount factors are calculated by these iterations process and are further analyzed to determine the best setting according to the predefined criteria which is minimum distance.

An iterative procedure involves the gradual discounting the original *BBA*s. The term m^{α^0,α^1} indicates *BBA* discounted value of α_1. Stop condition is distance d_{BPA}. Iteration that gives minimum distance is optimal number of iteration i_{opt}. Important here is that we can also find a source that differs mostly from the average value. It may be omitted from the calculations and it may be explored independently. The advantage of this described approach is that it does not need any meta-information about the reliability of sources.

4 Experimental Results

We tested the above described model on synthetic data. This is a simplified data. But they are based on data obtained from real-world recommendation systems which are the part of many e-shops. This is a simplified data. But they are based on data obtained from real-world recommendation systems which are the part of many e-shops. It means goods and services or their sets and respective recommendation expressed on the interval [0,1].



	{a}	{b}	{c}	{a,b}	{a,c}	{b,c}	{a, b. c}
m1	0.5	0	0.1	0.1	0.05	0.025	0.025
m2	0.52	0.12	0.08	0.05	0.1	0.06	0.07
m3	0.6	0.08	0.12	0.025	0.1	0.025	0.07
m4	0.2	0.1	0.6	0.05	0.025	0.025	0
m5	0.48	0.15	0.09	0.13	0.04	0.02	0.09
m6	0.45	0.21	0.11	0.09	0.06	0.05	0.03
m*	0.3989	0.1022	0.2762	0.0598	0.0785	0.0621	0.0223
m	0.5112	0.1899	0.1169	0.0895	0.049	0.0234	0.0201

Figure 1 The BBA set and the result of aggregation

The responses of various sources (recommenders) are represented by the values of belief functions in Figure 1. The six different sources are modeled (m_1–m_6). True value from recommendations has the same values as the values $m_1(\cdot)$. The value of $m^*(\cdot)$ is calculated using equation (4). The value of $m(\cdot)$ in the last but one row of the table is calculated according to the process outlined in the previous section. Source 4 m_4 is modeled as adversarial, because its reaction is opposite to the true value. The discount factor calculated for this source reaches the highest values. The table shows that discounting process overrides the impact of this source and as a result the result of the integration of information sources (recommendation) will be close to true value (m^*).

5 CONCLUSIONS

This article examines the problem of multiple recommenders which provide recommendation that are not entirely accurate. The problem concerns the use of model that is based on belief function theory and no additional information about the reliability of recommenders are known. Our approach provides an estimate of the true value from recommendations and also predicts the response of each recommender of the new instance. Experiments show that the proposed method outperforms several core values and leads to a performance close to the model trained with recommendation. There are many opportunities for further research. One possible direction is to extend our model with more cores learning. The aim is to choose an algorithm or a composite different covariance functions instead of fixing the combination in advance. Consequently, the algorithm may be difficult to learn fits recommender selecting multiple cores in data-dependent manner. In addition, it would be very useful to design efficient sampling methods for selection that instance and the response should be taught more. Our aim is to test further the described algorithm on further real data and deeper to verify the model described in this paper.

References

[1] Beranek, L., and Nydl, V.: The use of belief functions for the detection of internet auction fraud. In: *Proceedings of the 31st International Conference Mathematical Methods in Economics 2013*. Jihlava, College of Polytechnics Jihlava, 2013, 31 - 36.

[2] Beranek, L.: The usage of contextual discounting and opposition in determining the trustfulness of users in online auctions. *Journal of Theoretical and Applied Electronic Commerce Research* **7**(2012), 34-50.

[3] Carlson, J., Murphy, R., Christopher, S., and Casper, J.: Conflict metric as a measure of sensing quality. In: *ICRA 2005, IEEE International Conference on Robotics and Automation*. IEEE 2005, 2032-2039.

[4] Chen, S., Zhang, J., Chen, G., and Zhang, C.: What if the irresponsible teachers are dominating? In: *AAAI 2010, 24th Conference on Artificial Intelligence*. Atlanta, 2010, 419-424.
[5] Crammer, K., Kearns, M., and Wortman, J.: Learning from multiple sources. *Journal of Machine Learning Research* **9**(2008), 1757-1774.
[6] Daniel, M.: Conflicts within and between belief functions. In: *IPMU 2010, Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Dortmund, 2010, 696-705.
[7] Daniel, M.: Several notes on belief combination. In: *The Theory of Belief Functions Workshop*. Brest, ENSIETA, 2010, 48-53.
[8] Dawid, A. and Skene, A.: Maximum likelihood estimation of observer error-rates using the EM algorithm. *Applied Statistics* **28**(1979), 20-28.
[9] Dragoni, A. and Giorgini, P.: (2001). Revising beliefs received from multiple sources. In: *Frontiers in Belief Revision*, Applied Logic Series Volume 22. Springer, 2001, 429-442.
[10] Feng, N. and Yang, L.: Identification of the D-S evidence conflict based on multi-source pieces of information. In: *CSEE 2013, 2nd International Conference on Computer Science and Electronics Engineering*. Atlantis Press, 2013, 2725-2728.
[11] Han, X., Huang, X., and Eckert, C.: Learning from multiple observers with unknown expertise. In: *17th Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, LNAI 7818, 2013, 595-606.
[12] Hui, S. and Walter, S. Estimating the error rates of diagnostic tests. *Biometrics* **36**(1980), 167-171.
[13] Klein, J. and Colot, O.: Automatic discounting rate computation using a dissent criterion. In: *The Theory of Belief Functions Workshop*. Brest, 2010, 196-201.
[14] Liang-zhou, C., Wen-kang, S., Yong, D., and Zhen-fu, Z.: A new fusion approach based on distance of evidences. *Journal of Zhejiang Universtiy, SCIENCE A* **6**(2005), 476-482.
[15] Liu, W.: Analyzing the degree of conflict among belief functions. *Artificial Intelligence* **170**(2006), 909-924.
[16] Lysek, J., Stastny, J., and Motycka, A. Object recognition by means of evolved detector and classifier program. In: *Mendel 2012, 18th International Conference on Soft Computing*. Brno University of Technology, 2011, 456-462.
[17] Martin, A., Jouselme, A., and Osswald, C.: Conflict measure for the discounting operation on belief functions. In: *FUSION 2008, IEEE Int. Conf. on Information Fusion*. Springer, 2008, 1-8 .
[18] Murphy, C.: Combining belief functions with evidence conflicts. Decision Support Systems **29**(2000), 1-9.
[19] Raykar, V., Yu, S., Zhao, L., Valadez, G., Florin, C., Bogoni, L., and Moy, L.: Learning from crowds. *Journal of Machine Learning Research* **11**(2010), 1927-1322.
[20] Raykar, V., Zhao, L., Jerebko, A., Florin, C., Valadez, G., Bogoni, L., and Moy, L.: Supervised learning from multiple experts: Whom to trust when everyone lies a bit. In: *ICML 2009, 26th Annual International Conference on Machine Learning*. ACM, New York, 2009, 889-896.
[21] Ricquebourg, V., Delafosse, M., Delahoche, L., Marhic, B., Jolly-Desodt, A., and Menga, D.: Fault detection by combining redundant sensors : a conflict approach within the tbm framework. In: *COGIS 2007, Cognitive systems with interactive sensors*. Stanford, 2007, 26-28.
[22] Schafer, G.: *A mathematical theory of evidence*. Princeton University Press, Princeton, 1976.
[23] Schubert, J.: Conflict management in Dempster--Shafer theory using the degree of falsity. *International Journal of Approximate Reasoning* **52**(2011), 449-460.
[24] Skorpil, V., and Stastny, J.: Comparison Methods for Object Recognition. In: *Proceedings of the 13th WSEAS International Conference on Systems*. Rhodos, Greece, 2009. p. 607-610.
[25] Smets, P.: Analyzing the combination of conflicting belief functions. *Information Fusion* **8**(2007), 387-412.
[26] Smyth, P., Fayyad, U., Burl, M., Perona, P., and Baldi, P.: Inferring ground truth from subjective labelling of venus images. In: *NIPS 1994, Neural Information Processing Systems*. Wienna, 1994 1085-1092.
[27] Spiegelhalter, D. and Stovin, P.: An analysis of repeated biopsies following cardiac transplantation. *Statistics in Medicine* **2**(1983), 33-40.
[28] Stastny, J., Prochazka, D., Koubek, T., Landa, J.: Augmented Reality Usage for Prototyping Speed up. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **2**(2011), 353-360.
[29] Tubaishat, M. and Madria, S.: Sensor networks: an overview. *Potentials, IEEE* **22**(2003), 20-23.
[30] Whitehill, J., Ruvolo, P., Wu, T., Bergsma, J., and Movellan, J.: Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In: *NIPS 2009, Neural Information Processing Systems*. Vancouver, 2009, 2035-2043.
[31] Wu, O., Hu, W., and Gao, J.: Learning to rank under multiple annotators. In: *IJCAI 2011, 22nd International Joint Conference on Artificial Intelligence*. IJCAI/AAAI., London, 2011, 1571-1576
[32] Yan, Y., Rosales, R., Fung, G., Schmidt, M., Hermosillo, G., Bogoni, L., Moy, L., Dy, J., and Malvern, P.: Modeling annotator expertise: Learning when everybody knows a bit of something. In: *AISTATS 2010, 13rd International Conference on Artificial Intelligence and Statistics*. Chia Laguna Resort, Sardinia, 932-939.

Macroprudential experiment using the DSGE model of the Czech Republic with Housing sector

Milan Bouda[1]

Abstract. The recent financial crisis has made it evident - the necessity of introducing policies and regulations that adapt changes in the financial environment. In a fragile global economy, traditional measures have not seemed to be sufficient to avoid the crisis and have a fast and effective recovery. The paper describes the Small Open Economy Model of the Czech Republic with housing sector which is exactly derived in my finishing dissertation. DSGE model consists of households, firms, central bank and government. There are two types of households. The first type of consumers is able to access the capital markets and they can smooth consumption across time by buying or selling financial assets. The other type of household uses rule of thumb (ROT) consumption, spending all their income on consumption. Macroprudential experiment quantifies the impact of the level of Loan to Value (LTV) ratio to house prices and Gross Domestic Product. Czech National Bank may control this LTV ratio using macroprudential regulation which is obligatory for all banks in the Czech Republic.

Keywords: DSGE, Macroprudential policy, housing, LTV.

JEL classification: E58
AMS classification: 91B64

1 Introduction

The recent financial crisis has demonstrated the necessity of introducing policies and regulations that adapt to changes in the financial environment. In a fragile global economy, traditional macroprudential actions have not seemed to be sufficient to avoid the crisis and have a fast effective recovery. The crisis and its consequences have opened a real debate about the reforms that need to be made in the financial and regulatory banking system, and in the policy instruments that have to be used in order to avoid similar events. The new direction of policy interventions may be so-called macroprudential approach to mitigate the risk of the financial system as a whole that is the systemic risk. The term macroprudential refers to the use of prudential tools to explicitly promote the stability of the financial system in a global sense, not just the individual institutions (banks). The goal of this kind of regulation and supervision would be to avoid the transmission of financial shocks to the broader economy.

Financial crisis was caused largely by subprime mortgages which were provided in the U.S. These mortgages were normally approved to clients (borrowers) with lower credit ratings. A conventional mortgage is not offered because lender views the borrower as having a larger than average risk of defaulting on the loan. Lending institutions often charge interest on subprime mortgages at a rate that is higher than a conventional mortgage in order to compensate them for carrying more risk. These borrowers are very sensitive to any economic downturn. When U.S. economy started to fall into recession then exactly this kind borrowers became defaults. Next, lenders (banks and other financial institutions) repossessed collateral (house) which was subject of subprime mortgage. When workout departments of all lenders start to sell repossessed collaterals then house prices start to decline. This has impact also on non-defaulted clients and market price of their houses. In a very extreme case that market house prices are halved compare to value valid to time of application, the borrowers lose motivation to pay the principal and borrowers let financial institution to repossess their house. This situation happened in the U.S. after financial crisis.

[1]University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, xboum00@vse.cz

2 Model

The macroprudential experiment can not be performed without any econometric model. Thus, the Small Open Economy (SOE) DSGE model of the Czech Republic with housing sector must be presented.

The model used in this paper is broadly inspired by [2] (APV). Their model was constructed as a closed economy model. The modified version of APV model incorporates as open economy extension and also government. This inclusion of both foreign sector and government makes the model more applicable to a small open economy such as Czech Republic. The key open economy features are as follows. The production function contains an imported intermediate good. The goods producing sector sells to foreign consumers in addition to home consumers. A proportion of consumers are able to access foreign capital markets. The model is extended by a specific fiscal authority called government. Intuitively, one would expect that the stimulatory effect of government spending on the economy would naturally increase the level of economic activity.

Following the framework used by APV, there are two types of households. The first type of consumers is able to access the capital markets and they can smooth consumption across time by buying or selling financial assets. These households follow the permanent income hypothesis (PIH). The other type of household uses rule of thumb (ROT) consumption, spending all their income on consumption. ROT consumers are effectively completely credit constrained as they do not have any access to the credit markets. This dual differentiation of consumers is based on [3].

Unlike the APV model, in this study ROT households are further characterized by the fact that they do not own any housing assets. This guarantees consistency so that households which do not have access to the credit markets to smooth consumption are also unable to purchase a house (it means that mortgage would not be granted to them). Given that residential mortgage lending institutions would be unwilling to extend credit to potential borrowers who have no assets, not to mention insufficient funds for a deposit on a house, this is not only reasonable but also arguably more realistic than positing credit-constrained consumers repeatedly accessing the mortgage market. Both types of consumers purchase goods from firms each period, receive wage income from labor supplied to firms and pay rental to the homeowners.

PIH households are divided into two complementary components: a homeowner and a consumer. The homeowner transacts in the housing market each period, selling the housing stock and purchasing the stock anew. Against the net worth of housing stock, the homeowner borrows to meet any shortfall between the price of the housing stock bought at the end of the period and the price realized on sale of the existing housing stock. Net worth is defined as the value of the housing stock less outstanding debt and less any dividends paid to consumers. This dividend is the mechanism by which the housing equity withdrawal is captured. Homeowners also charge a rental fee to consumers. Thus the housing stock is completely owner by the PIH consumers and the ROT consumers pay rental to their PIH landlords.

Firms are monopolistically competitive and produce a continuum of consumer goods. Each period they hire labor from households and also purchase an intermediate input from abroad. These imports are used by firms each period and capital is effectively assumed to be constant. The output of firms is consumed by either household or government, exported or used to produce additional housing stock. The conversion of consumer goods to housing stock is assumed to follows q-type investment theory[1].

The monetary authority has a Taylor rule reaction function (with lagged inflation and the output gap as indicators of inflationary pressure) and uses the nominal interest rate as its lever subject to a smoothing parameter.

The government collects lump sum taxes from consumers and purchases consumer goods. The difference between these two is either funded through the sale of government bonds or, where taxes exceed expenditure, is used to retire debt. For simplicity government expenditure does not impact households directly (in other models it is done in the form of transfer) but rather is a source of final demand for consumer goods and thence labor demand and imports. Following [4], fiscal policy is modeled as the combination of exogenous government spending, government debt and lump sum taxes.

Figure 1 shows the structure and dynamics of SOE DSGE model with housing. Each agent mentioned

[1] [8] writes "One, the numerator, is the market valuation: the going price in the market for exchanging existing assets. The other, the denominator, is the replacement or reproduction cost: the price in the market for the newly produced commodities. We believe that this ratio has considerable macroeconomic significance and usefulness, as the nexus between financial markets and markets for goods and services." The formula may be for purposes of this study rewritten as follows $q = \frac{\text{value of housing stock}}{\text{net worth}}$, where q represents the real house prices.

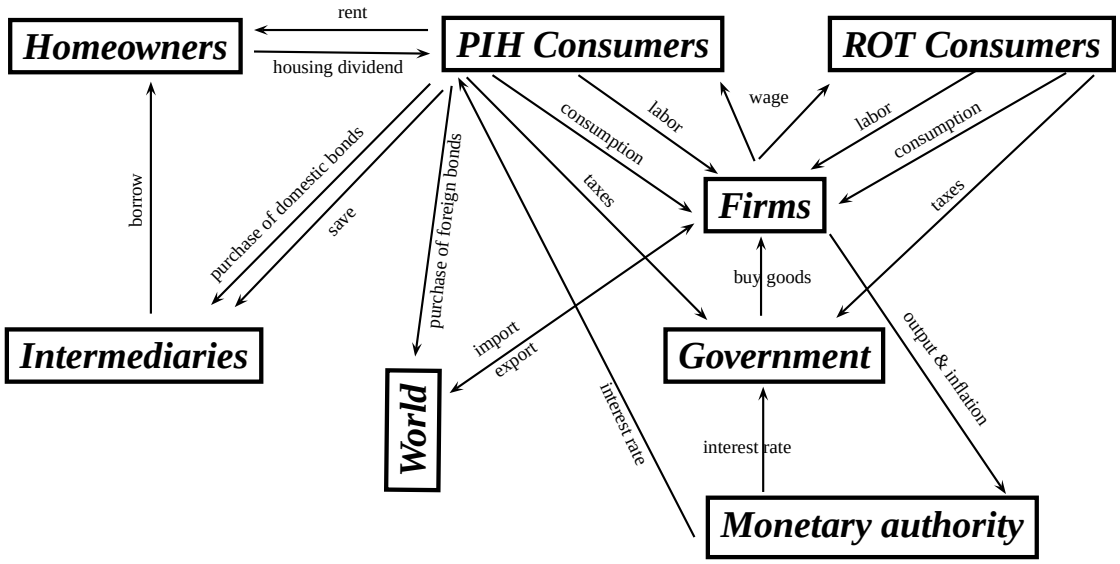


Figure 1 The structure of the SOE DSGE model with housing

above maximizes its utility function. The goal is to obtain a set of first order conditions characterizing the equilibrium of this SOE model.

Model contains 37 equations, 37 variables, 5 observed variables, 5 stochastic shocks and 29 structural parameters. Such an extensive model can not be presented in this paper. The complete derivation, documentation and scripts may be found in my coming dissertation thesis. In case of interest I can provide complete derivation of my DSGE model as well as Dynare code.

2.1 Data and Bayesian estimation

The presence of five stochastic shocks allows verification of how this DSGE model fits observed data. Each stochastic shock enables application of one observed variable. Thus the following five observed variables are introduced: The Gross Domestic Product (GDP), real house prices, nominal exchange rate, inflation and nominal interest rate. The time series of these variables are taken from International Monetary Fund (IMF), database of the Czech Statistical Office (CSO) and database of the Czech National Bank (ARAD). Time series are on quarterly basis and due to limitation of house prices data I use the period from Q1 2006 to Q3 2013 which means 31 quarterly observations.

Real GDP is taken from the database of the IMF. GDP is in constant prices of the year 2005 and data are seasonally adjusted. Next, GDP is transformed to the YoY (Year over Year) percentage changes.

Calculation of inflation is based on CPI (Consumer Price Index) which is in real terms (year 2005 = 100). CPI is taken from IMF database. This index is transformed to the YoY percentage changes which represent the inflation.

Nominal exchange rate is represented by the Nominal Effective Exchange Rate of the Czech Republic which is taken from database IMF. This is an index (year 2005 = 100). YoY percentage changes of this index may be interpreted as changes of competitiveness.

Nominal interest rate is represented by Prague Inter Bank Offered Rate (PRIBOR). Inter Bank rate is the rate of interest charged on short-term loans made between banks. Banks borrow and lend money in the interbank market in order to manage liquidity and meet the requirements placed on them. The quarterly data is used in this study thus I decided to apply 3 months PRIBOR rate which is published in ARAD.

Real house prices are taken from the database of the CSO. House prices are in index (year 2005 = 100). House prices are represented by realized prices of flats. This index is then transformed to the YoY percentage changes. Next, these YoY % changes (housing price inflation) minus inflation (CPI inflation) represent the approximation of the real house prices.

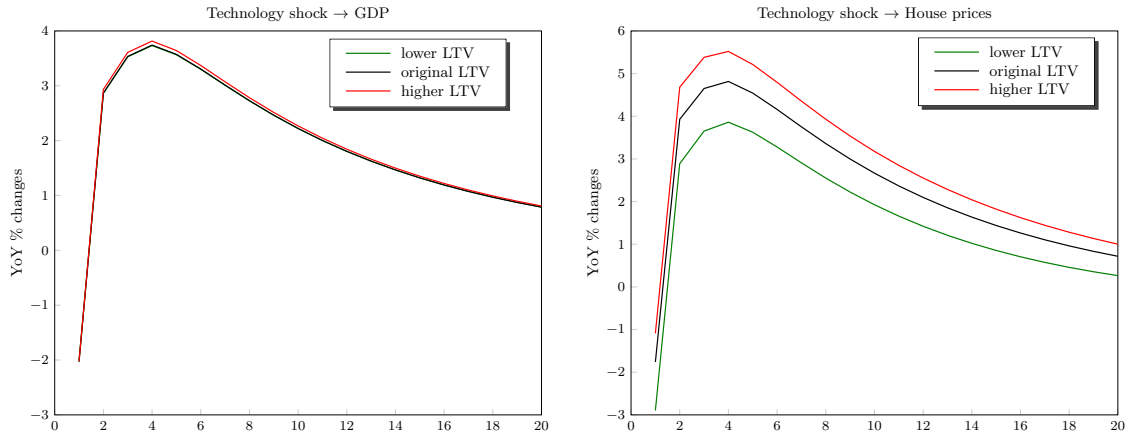


Figure 2 Impulse response functions of technology shock

In order to perform macroprudential experiment all parameters of DSGE models must be estimated. Estimation of all structural parameters is performed using Maximum Likelihood through iterative Kalman-filtering processes. As posterior distributions often have unknown distribution patterns, numerical random sample-generating techniques (such as the Metropolis-Hastings algorithm) are usually involved in DSGE estimation. Model was estimated using the industry-standard approach: Dynare is used, a freeware add-on to the MATLAB software. Details on the Bayesian methodology used for our estimations may be observed from [7] and technical aspects of Dynare implementation are available from [1] and [5].

3 Macroprudential experiment

The goal is to investigate what are the impacts of introduction less strict and stricter Loan to Value[2] (LTV) rule. Introduction of lower (stricter) LTV ratio should mitigate the risk which is connected with default of counterparty. On the other hand, higher (less strict) LTV ratio supports economy in good times but also tends to deepen crisis in bad times.

This model does not contain such thing as LTV ratio. On the other hand, model comprises the net worth ratio

$$\phi = \frac{N}{qh}, \tag{1}$$

which can be perceived as $1 - LTV$, but not on household level but on economy level. N is the net worth which is calculated as housing stock less outstanding debt less dividends paid to consumers, q represents the real house prices and h is the housing stock. The economic interpretation is as follows. The lower net worth ratio means that current level of housing stock is mainly financed by housing loans. On the other hand, the higher net worth ratio means that current level of housing stock is not financed by housing loans and is almost fully owned by PIH households.

Following 1 one obtain

$$LTV = 1 - \phi = 1 - \frac{N}{qh}.$$

Now, the LTV ratio has the usual interpretation. The lower LTV means that there is a high level of collateralization in our economy. On the other hand, higher LTV means that majority of housing stock is covered by housing loans.

The experiment performed in this chapter is as follows. Figures 2, 3, 4, 5 and 6 show behavior of impulse response functions to technology shock, government spending shock, foreign interest rate shock, domestic interest rate shock and foreign demand shock. The original LTV is in SOE model set to 0.3. The higher LTV is set to 0.4 and lower LTV is set to 0.2.

[2]The Loan to Value (LTV) ratio is a financial term used by lenders to express the ratio of Mortgage Amount to Appraised Value of the Property.

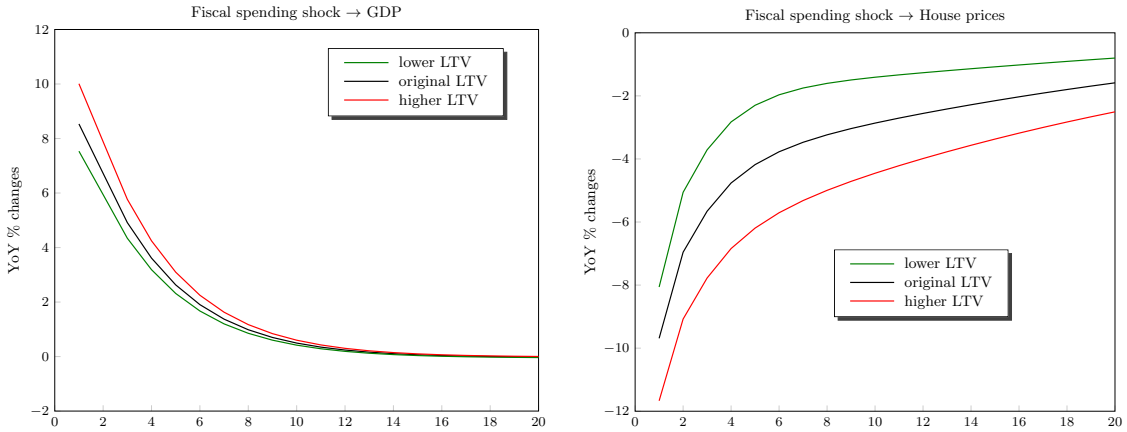


Figure 3 Impulse response functions of government spending shock

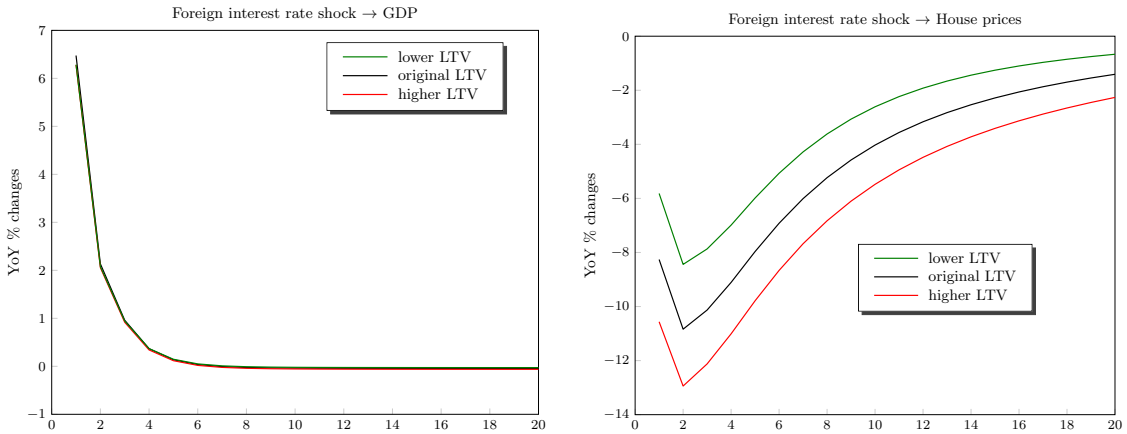


Figure 4 Impulse response functions of foreign interest rate shock

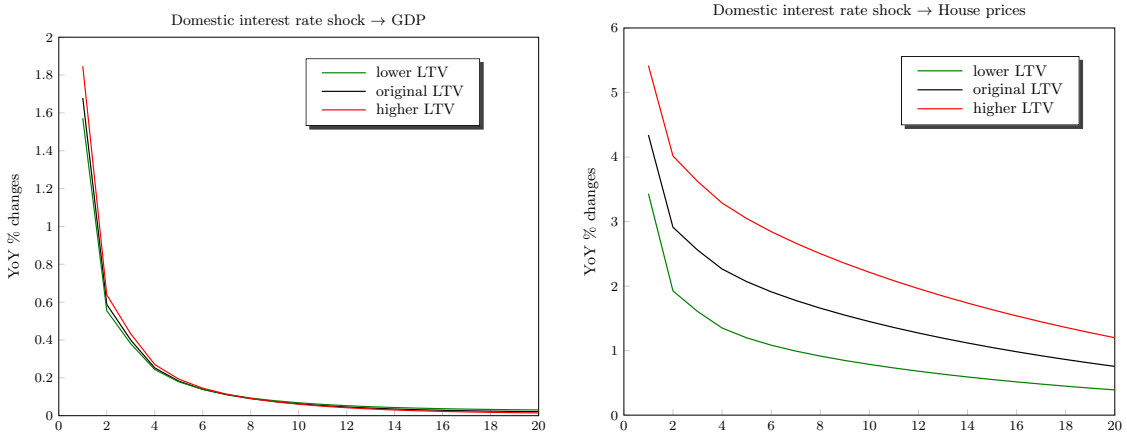


Figure 5 Impulse response functions of domestic interest rate shock

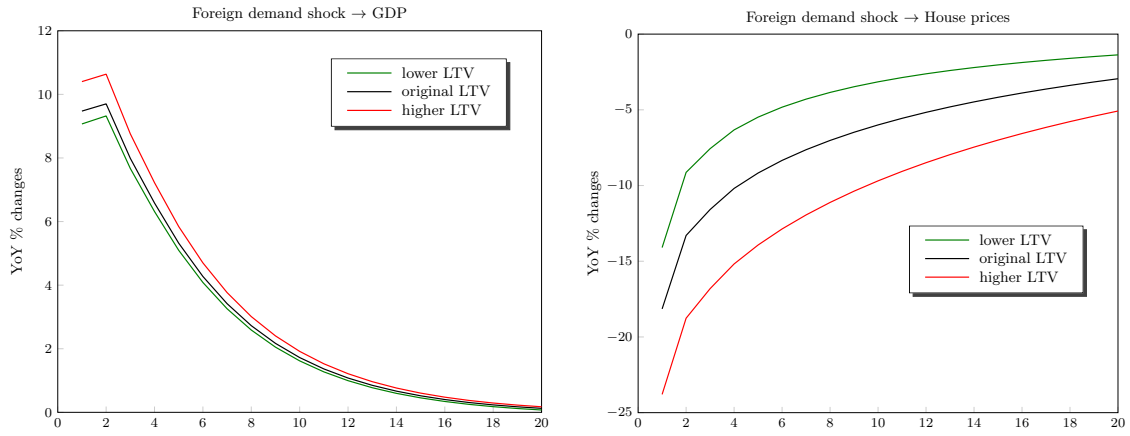


Figure 6 Impulse response functions of foreign domestic demand shock

4 Conclusion

First, the problematic of macroprudential policy is introduced and SOE DSGE model with housing sector is shortly described as well as data which are used in Bayesian estimation of this model. Second, the macroprudential experiment is performed using this model. One may see that behavior of GDP is almost the same for all three levels of LTV. In contrast with GDP one can see that house prices are more sensitive to level of LTV. This reality is in accordance with former research of [6]. The results are very conclusive. This means that higher LTV ratio in the economy causes the higher and more volatile house prices than in case of original or lower LTV. This fact is observed for all stochastic shocks in the model. In order to keep house prices less volatile it is recommended to implement rather more restrictive LTV. This recommendation is based on fact that investors prefer stable economic environment.

Acknowledgements

The research was supported by Grant No. VŠE IGA F4/24/2014, FIS, University of Economics, Prague.

References

[1] Adjemian, S., Bastani, H., Karame, F., Juillard, M., Maih, J., Mihoubi, F., Perendia, G., Ratto, M., Villemot, S.: Dynare: Reference Manual Version 4. *CEPREMAP*, (2011).

[2] Aoki, K., Proudman, J., Vlieghe, G.: Houses as collateral: has the link between house prices and consumption in the U.K. changed?. *Economic Policy Review*, May (2002), 163–177.

[3] Campbell, J. Y., Mankiw, N. G.: Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence. *NBER Macroeconomics Annual*, Volume 4, (1989), 185–246.

[4] Galí, J., Gertler, M.: Macroeconomic Modeling for Monetary Policy Evaluation. *NBER Working Papers*, Number 13542, (2007).

[5] Griffoli, T. M.: DYNARE User Guide. *Dynare project*, (2013).

[6] Iacoviello, M., Neri, S.: Housing Market Spillovers: Evidence from an Estimated DSGE Model. *American Economic Journal: Macroeconomics*, Volume 2(2), (2010), 125–164.

[7] Koop,G.: *Bayesian Econometrics*. John Wiley & Sons Inc., 2003.

[8] Tobin, J.: A General Equilibrium Approach to Monetary Theory. *Journal of Money, Credit and Banking*, Volume 1(1), (1989), 15–29.

Sample approximation techniques for DEA-risk efficiency tests

Martin Branda[1]

Abstract. The new class of data envelopment analysis (DEA) models with diversification suitable for financial applications has gained a special attention in recent years. These models take into account possible dependencies between the considered asset returns which were not considered by the standard DEA models. Several solution methods have been proposed for the resulting nonlinear problems, however, these methods are restricted to discrete distributions only. In this paper, we consider a multivariate continuous distribution of the random returns and use the sample approximation technique to solve the DEA problems approximately. We focus on the quality of the approximation. Moreover, we introduce a one-step procedure where inputs for a benchmark need not to be computed before solving the DEA model. The new methods are applied to access efficiency of selected assets from US stock market.

Keywords: Data envelopment analysis, efficiency, multivariate skew-normal distribution, sample approximation technique, one-step procedure.

JEL classification: C44
AMS classification: 90C15

1 Introduction

We deal with the problem of efficiency of investment opportunities, i.e. assets or portfolios, available on financial markets. There are several approaches to this issue based on utility functions, mean-risk models, stochastic dominance and Data Envelopment Analysis (DEA). In this paper, we focus on DEA models suitable for financial applications. DEA with diversification was introduced to remove the drawback of the standard DEA models which do not take into account dependencies between the considered investment opportunities when risk measures are used as inputs. The first diversification-consistent DEA model was introduced by Briec et al. [8] who investigated mean-variance efficiency (Markowitz [14]) within the DEA framework. This approach was further extended by Joro and Na [11] and Briec et al. [9] who added skewness to the models as an input. Lozano and Gutiérrez [13] dealt with simple DEA models with diversification that were shown to provide a necessary condition for efficiency with respect to second-order stochastic dominance (SSD). Branda and Kopa [6, 7] continued in this research direction and found DEA models which are even equivalent to SSD efficiency tests. Recently, general diversification-consistent DEA models were introduced based on the classes of coherent risk measures, cf. Lamb and Tee [12], and general deviation measures, see Branda [3, 4]. It can be shown that for convex deviation measures and concave return measures the diversification-consistent models are stronger than the traditional ones, see Branda [3]. Thus, traditional models overestimate efficiency scores of investment opportunities.

In this paper, we focus on a particular diversification-consistent DEA models where the weighted mean absolute deviations from α-quantile, known also as CVaR deviations, are used to access risk of investment opportunities, i.e. they serve as the inputs. Expected return is considered as the only output. This model was introduced and further developed by Branda [3, 4]. However, the formulations were always restricted to finite discrete distribution of random returns, whereas continuous distribution is much more desirable for their modeling. The resulting problems are then hardly solvable directly and some approximation technique is necessary. We employ the sample approximation technique to

[1]Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 83, Prague 186 75, Czech Republic
tel.: +420 221 913 404, fax: +420 222 323 316, branda@karlin.mff.cuni.cz

approximate the continuous distribution by independent Monte-Carlo samples which are used as discrete distributions with equiprobable realizations. Diversification-consistent models are usually performed in two steps when the inputs and outputs for a benchmark are evaluated first and then the efficiency score is computed. We will avoid this two-step procedure by introducing a one-step formulation of the models, which is helpful in solving of many sample approximated problems.

This paper is organized as follows. In Section 2, we introduce the diversification consistent DEA models. We propose a one-step formulation of the DEA model with CVaR deviations in Section 3. In Section 4, we apply the sample approximation technique to access efficiency of selected assets from US stock market. Section 5 concludes the paper.

2 DEA-risk models with diversification

In this section, we formulate a general diversification consistent DEA model. General deviation measures are used to access risk of the investment opportunities and serve as the inputs, which are preferred to have lower values. Return measures are employed in general as the outputs and are maximized by the investors. Both deviation and return measures fulfill the axioms given below.

Let $\mathcal{X}$ be a set of random returns of available investment opportunities corresponding to a single asset or to a portfolio consisting of many assets. We consider n assets and denote R_i the return of i-th asset which is a nondegenerated real random variable defined on the probability space $(\Omega, \mathcal{A}, P)$. We will use the set with no short sales enabling full diversification: $\mathcal{X} = \{\sum_{i=1}^n R_i x_i : \sum_{i=1}^n x_i = 1, x_i \geq 0\}$. Other choices of the set are also possible, e.g. with limited number of assets, allowing short sales, borrowing or including proportional and fixed transaction costs, cf. Branda [3], Ding et al. [10].

A functional $\mathcal{D} : \mathcal{L}_2(\Omega) \to [0, \infty]$ is called the general deviation measure (Rockafellar et al. [16]) if they satisfy:
(D1) translation invariance: $\mathcal{D}(X + C) = \mathcal{D}(X)$ for all X and constants C,
(D2) positive homogeneity: $\mathcal{D}(0) = 0$, and $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$ for all X and all $\lambda > 0$,
(D3) subadditivity: $\mathcal{D}(X + Y) \leq \mathcal{D}(X) + \mathcal{D}(Y)$ for all X and Y,
(D4) nonnegativity: $\mathcal{D}(X) \geq 0$ for all X, with $\mathcal{D}(X) > 0$ for nonconstant X.
Note that the axioms (D2) and (D3) imply convexity. The examples are standard deviation, semideviations and CVaR deviation. It corresponds to the weighted mean absolute deviation from the $(1 - \alpha)$-th quantile, see Ogryczak and Ruszczynski [15], and is defined as

$$CVaR_\alpha^{dev}(X) = \min_{\xi \in \mathbb{R}} \frac{1}{1-\alpha} \mathbb{E}[\max\{(1-\alpha)(X - \xi), \alpha(\xi - X)\}], \tag{1}$$

with the minimum attained at any $(1 - \alpha)$-th quantile of the distribution X.

A functional $\mathcal{E} : \mathcal{L}_p(\Omega) \to (-\infty, \infty]$ is called the return measure (Lamb and Tee [12]) if it satisfies:
(E1) translation equivariance: $\mathcal{E}(X + C) = \mathcal{E}(X) + C$ for all X and constants C,
(E2) positive homogeneity: $\mathcal{E}(0) = 0$, and $\mathcal{E}(\lambda X) = \lambda \mathcal{E}(X)$ for all X and all $\lambda > 0$,
(E3) superadditivity: $\mathcal{E}(X + Y) \geq \mathcal{E}(X) + \mathcal{E}(Y)$ for all X and Y,
(E4) monotonicity: $\mathcal{E}(X) \geq \mathcal{E}(Y)$ when $X \geq Y$.
The space $\mathcal{L}_p(\Omega)$ is selected so as the measures are finite, usually $p = 1$ is sufficient. Note that the axioms (E2) and (E3) imply concavity of the functional. It is obvious that expectation fulfills the axioms. Moreover, coherent risk measures multiplied by a negative constant can be used as return functionals too, see Lamb and Tee [12].

Traditional mean-deviation efficiency can be defined as follows: an investment opportunity is efficient if there is no other investment opportunity with higher or equal expected return and lower of equal deviation with at least one inequality strict. This approach to efficiency is extended by the DEA models, where K deviation measures and J return measures are employed at the same time. The input oriented model can be formulated as follows

$$\begin{aligned} \theta(R_0) = \min_{\theta, X} \theta & \\ \text{s.t. } \mathcal{E}_j(X) &\geq \mathcal{E}_j(R_0),\ j = 1, \ldots, J, \\ \mathcal{D}_k(X) &\leq \theta \cdot \mathcal{D}_k(R_0),\ k = 1, \ldots, K, \\ X &\in \mathcal{X}. \end{aligned}$$

We say that an investment opportunity is efficient if the optimal value is equal to 1, otherwise we say that it is inefficient. Since the general deviation measures are convex and the return measures are assumed to be concave, we obtain a convex programming problem if the set $\mathcal{X}$ is convex. The theory of diversification-consistent DEA models was recently developed by Lamb and Tee [12] and Branda [3, 4].

3 Sample approximations for DEA-risk models

In this section, we deal with the sample approximation technique, see Shapiro et al. [17], Branda [2, 5] for an introduction and general results. Moreover, we introduce the one-step procedure for solving the diversification-consistent model. Let the returns have a multivariate continuous distribution and let $(r_{1s}, \ldots, r_{ns})$, $s = 1, \ldots, S$, be an independent Monte-Carlo sample, i.e. i.i.d. random vectors with equal probabilities $1/S$. CVaR deviation on level α can be then computed as

$$CVaR_\alpha^{dev}\left(\sum_{i=1}^n R_i x_i\right) = \min_{\xi\in\mathbb{R}} \frac{1}{S}\sum_{s=1}^S \max\left\{(\sum_{i=1}^n x_i r_{is} - \xi), \frac{\alpha}{1-\alpha}(\xi - \sum_{i=1}^n x_i r_{is})\right\}.$$

We give a useful reformulation in the following proposition.

Proposition 1. *Under a discrete distribution of returns, the CVaR deviation can be computed as the optimal value of the following linear programming problem*

$$\begin{aligned} CVaR_\alpha^{dev}(R_0) = \max_{y_s, z_s} \sum_{s=1}^S r_{0,s} y_s - \frac{\alpha}{1-\alpha}\sum_{s=1}^S r_{0,s} z_s \\ \sum_{s=1}^S y_s - \frac{\alpha}{1-\alpha}\sum_{s=1}^S z_s &= 0, \\ y_s + z_s &= \frac{1}{S}, \\ y_s, z_s &\geq 0. \end{aligned} \tag{2}$$

Proof. CVaR deviation can be expresses as optimal value of the following program

$$CVaR_\alpha^{dev}(R_0) = \min_{\xi\in\mathbb{R}} \frac{1}{S}\sum_{s=1}^S \max\left\{r_{0,s} - \xi, \frac{\alpha}{1-\alpha}(\xi - r_{0,s})\right\},$$

which can be reformulated as a linear programming problem

$$\begin{aligned} \min_{\xi, u_s} \frac{1}{S}\sum_{s=1}^S u_s \\ u_s + \xi &\geq r_{0,s}, \\ u_s - \frac{\alpha}{1-\alpha}\xi &\geq -\frac{\alpha}{1-\alpha} r_{0,s}. \end{aligned}$$

Then linear programming duality can be employed resulting into (2). □

We consider $CVaR_\alpha^{dev}$ for various levels $\alpha_k \in (0,1), k = 1, \ldots, K$, as the inputs and the expectation as an output, i.e. $J = 1$ and $\mathcal{E}_1(X) = \mathbb{E}X$. Standard diversification-consistent model is computed in two steps. Inputs and outputs of the benchmark have to be evaluated first, then the DEA-risk model is solved. In the following preposition we give a one-step formulation, where CVaR deviations are evaluated during the model computation.

Proposition 2. *Under the discrete distribution of returns, the one-step DEA model without precompu-*

tation of CVaR deviations can be formulated as the following linear programming problem

$$\begin{aligned}
\theta^S(R_0) = \min_{\theta,x_i,u_{sk},\xi_k,\tilde{y}_{sk},\tilde{z}_{sk}} & \theta \\
\text{s.t. } \frac{1}{S}\sum_{i=1}^{n}\sum_{s=1}^{S} r_{i,s}x_i &\geq \frac{1}{S}\sum_{s=1}^{S} r_{0,s}, \\
\frac{1}{S}\sum_{s=1}^{S} u_{sk} &\leq \sum_{s=1}^{S} r_{0,s}y_{sk} - \frac{\alpha}{1-\alpha}\sum_{s=1}^{S} r_{0,s}z_{sk} \\
\sum_{i=1}^{n} x_i r_{is} - \xi_k &\leq u_{sk},\ s=1,\ldots,S, k=1,\ldots,K, \\
\frac{\alpha_k}{1-\alpha_k}\left(\xi_k - \sum_{i=1}^{n} x_i r_{is}\right) &\leq u_{sk},\ s=1,\ldots,S, k=1,\ldots,K, \\
\sum_{s=1}^{S} y_{sk} - \frac{\alpha_k}{1-\alpha_k}\sum_{s=1}^{S} z_{sk} &= 0, \\
y_{sk} + z_{sk} &= \frac{\theta}{S},\ s=1,\ldots,S, k=1,\ldots,K, \\
y_{sk}, z_{sk} &\geq 0,\ 0 \leq \theta \leq 1, \\
\sum_{i=1}^{n} x_i &= 1,\ x_i \geq 0,\ i=1,\ldots,n.
\end{aligned} \tag{3}$$

Proof. In Branda [3], it was shown that the problem with CVaR deviations leads to the following linear programming problem:

$$\begin{aligned}
\theta^S(R_0) = \min_{\theta,x_i,u_{sk},\xi_k} & \theta \\
\text{s.t. } \frac{1}{S}\sum_{i=1}^{n}\sum_{s=1}^{S} r_{i,s}x_i &\geq \frac{1}{S}\sum_{s=1}^{S} r_{0,s}, \\
\frac{1}{S}\sum_{s=1}^{S} u_{sk} &\leq \theta \cdot CVaR^{dev}_{\alpha_k}(R_0),\ k=1,\ldots,K, \\
u_{sk} &\geq \left(\sum_{i=1}^{n} x_i r_{is} - \xi_k\right),\ s=1,\ldots,S, k=1,\ldots,K, \\
u_{sk} &\geq \frac{\alpha_k}{1-\alpha_k}\left(\xi_k - \sum_{i=1}^{n} x_i r_{is}\right),\ s=1,\ldots,S, k=1,\ldots,K, \\
\sum_{i=1}^{n} x_i &= 1,\ x_i \geq 0,\ i=1,\ldots,n.
\end{aligned}$$

However, CVaR deviations for the benchmark have to be computed before the model is evaluated. If we use (2), we obtain the following nonlinear constraints

$$\frac{1}{S}\sum_{s=1}^{S} u_{sk} \leq \theta\left(\sum_{s=1}^{S} r_{0,s}y_{sk} - \frac{\alpha_k}{1-\alpha_k}\sum_{s=1}^{S} r_{0,s}z_{sk}\right).$$

By substituting $\tilde{y}_{sk} = \theta y_{sk}$ and $\tilde{z}_{sk} = \theta z_{sk}$, we obtain a linear system

$$\begin{aligned}
\frac{1}{S}\sum_{s=1}^{S} u_{sk} &\leq \max_{\xi,u}\sum_{s=1}^{S} r_{0,s}\tilde{y}_{sk} - \frac{\alpha_k}{1-\alpha_k}\sum_{s=1}^{S} r_{0,s}\tilde{z}_{sk} \\
\sum_{s=1}^{S} \tilde{y}_{sk} - \frac{\alpha_k}{1-\alpha_k}\sum_{s=1}^{S} \tilde{z}_{sk} &= 0, \\
\tilde{y}_{sk} + \tilde{z}_{sk} &= \frac{\theta}{S},\ s=1,\ldots,S, k=1,\ldots,K, \\
\tilde{y}_{sk}, \tilde{z}_{sk} &\geq 0,\ 0 \leq \theta \leq 1.
\end{aligned}$$

Note that it always holds $\theta > 0$, since otherwise the benchmark would be constant, which is in contradiction with our assumptions. This finishes the proof. □

4 Numerical study

In this section, we employ the diversification-consistent DEA models together with the sample approximation technique to access efficiency of selected US assets observed monthly from January 2009 to December 2013. In particular, we consider Boeing (BA), Coca-Cola (KO), JPMorgan Chase (JPM), Oracle (ORCL), Microsoft (MSFT), Nike (NKE), Intel (INTC), Apple (AAPL). The historical prices were downloaded using the function *FinancialData* available in Wolfram Mathematica 9. We employ multivariate skew-normal distribution. A n-dimensional random vector is said to have a multivariate skew-normal distribution if it has density function

$$f_n(z) = 2\phi_n(z, \Omega)\Phi(\beta^T z),\ z \in \mathbb{R}^n,$$

where $\phi_n(z, \Omega)$ is the n-dimensional normal density with zero mean and correlation matrix Ω, Φ is the N(0, 1) distribution function, and β is a n-dimensional vector, see Azzalini and Dalla Valle [1] for details. The parameters of the returns distribution were estimated using the R packages *fAssets* and *sn*. The same packages were also used to Monte-Carlo simulation. To test performance of the sample approximation technique, 100 independent samples of 1000 return realizations were simulated.

In the model (3), we consider $K = 4$ CVaR deviations for the levels $\alpha_k \in \{0.75, 0.9, 0.95, 0.99\}$ as the inputs and the expectation as an output, i.e. $J = 1$ and $\mathcal{E}_1(X) = \mathbb{E}X$. We solved the DEA problems by the CPLEX 12.1 solver using the modeling system GAMS 23.2. It took approximately 22 minutes of CPU time to solve the 800 approximated linear programming problems. All computations were performed on PC with Intel Core i7 2.90 GHz CPU, 8 GB RAM and 64-bit Windows 7 Professional operational system. Note that for-cycles together with an automatic data replacement available through the algebraic language in GAMS were used. Descriptive statistics of the resulting efficiency scores and confidence interval based on asymptotic normality can be found in Table 1. If we take the ranking according to the mean score as the reference one, we can investigate the difference in ranking using different samples. The results can be found in Table 2. We can see that the ranking is stable for the best two and the worst two funds.

	BA	KO	JPM	ORCL	MSFT	NKE	INTC	AAPL
Mean	0.597	0.884	0.475	0.545	0.627	0.706	0.673	1.000
Standard deviation	0.036	0.020	0.035	0.025	0.029	0.047	0.032	0.000
Minimum	0.505	0.843	0.417	0.473	0.560	0.582	0.605	1.000
Maximum	0.691	0.936	0.598	0.636	0.697	0.829	0.775	1.000
Confidence int. l.b.	0.590	0.880	0.468	0.540	0.621	0.697	0.666	1.000
Confidence. int. u.b.	0.604	0.888	0.482	0.550	0.633	0.716	0.679	1.000

Table 1 Efficiency scores – descriptive statistics and 95% confidence intervals

Ranking	6	2	8	7	5	3	4	1
Difference	BA	KO	JPM	ORCL	MSFT	NKE	INTC	AAPL
-2	6	0	1	0	0	0	0	0
-1	20	0	3	10	10	0	25	0
0	66	100	96	86	67	75	67	100
1	8	0	0	4	20	17	8	0
2	0	0	0	0	3	5	0	0
3	0	0	0	0	0	3	0	0

Table 2 Efficiency scores – differences in ranking for 100 samples

5 Conclusions

In this paper, we have proposed a one-stage procedure for solving the diversification-consistent DEA model with CVaR deviations. Moreover, we have employed the sample approximation technique to solve

the problem with the multivariate skew-normal distribution of the returns. The numerical results have confirmed good performance of the approximation.

Acknowledgements

This work was supported by the Czech Science Foundation under the Grant GP13-03749P.

References

[1] Azzalini, A., Dalla Valle, A.: The multivariate skew-normal distribution. *Biometrika* **83** (4) (1996), 715–726.

[2] Branda, M.: Stochastic programming problems with generalized integrated chance constraints. *Optimization* **61** (8) (2012), 949–968.

[3] Branda, M.: Diversification-consistent data envelopment analysis with general deviation measures. *European Journal of Operational Research* **226** (3) (2013), 626–635.

[4] Branda, M.: Reformulations of input-output oriented DEA tests with diversification. *Operations Research Letters* **41** (5) (2013), 516–520.

[5] Branda, M.: Sample approximation technique for mixed-integer stochastic programming problems with expected value constraints. *Optimization Letters* **8** (3) (2014), 861–875.

[6] Branda, M., Kopa, M.: DEA-risk efficiency and stochastic dominance efficiency of stock indices. *Czech Journal of Economics and Finance* **62** (2) (2012), 106–124.

[7] Branda, M., Kopa, M.: On relations between DEA-risk models and stochastic dominance efficiency tests. *Central European Journal of Operations Research* **22** (1) (2014), 13–35.

[8] Briec, W., Kerstens, K., Lesourd, J.-B.: Single period Markowitz portfolio selection, performance gauging and duality: a variation on the Luenberger shortage function. *Journal of Optimization Theory and Applications* **120** (1) (2004), 1–27.

[9] Briec, W., Kerstens, K., Jokung, O.: Mean-variance-skewness portfolio performance gauging: a general shortage function and dual approach. *Management Science* **53** (2007), 135–149.

[10] Ding, H., Zhou, Z., Xiao, H., Ma, C., Liu, W.: Performance Evaluation of Portfolios with Margin Requirements, *Mathematical Problems in Engineering*, vol. 2014, Article ID 618706, 8 pages, 2014. doi:10.1155/2014/618706

[11] Joro, T., Na, P.: Portfolio performance evaluation in a mean–variance–skewness framework. *European Journal of Operational Research* **175** (2006), 446–461.

[12] Lamb, J.D., Tee, K-H.: Data envelopment analysis models of investment funds. *European Journal of Operational Research* **216** (3) (2012), 687–696.

[13] Lozano, S., Gutiérrez, E.: Data envelopment analysis of mutual funds based on second-order stochastic dominance. *European Journal of Operational Research* **189** (2008), 230–244.

[14] Markowitz, H.M.: Portfolio selection. *The Journal of Finance* **7** (1) (1952), 77–91.

[15] Ogryczak, W., Ruszczynski, A.: Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* **13** (2002), 60–78.

[16] Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Generalized deviations in risk analysis. *Finance and Stochastics* **10** (2006), 51–74.

[17] Shapiro, A., Dentcheva, D., Ruszczyski, A.: *Lectures on stochastic programming: Modeling and theory*. SIAM, Philadelphia, 2009.

Fuzzy Approach to Risk Appetite in Project Management

Brožová Helena[1], Bartoška Jan[2], Šubrt Tomáš[3]

Abstract. Risk management is an important part of the project management. Companies undertake increasingly complex and ambitious projects, which must be executed successfully under uncertainty and risky environment. Therefore, the project threats assessment is crucial. Unfortunately, it is often relative to the projects. An acceptable risk to one project might be unacceptable to another. From this reason, it is difficult to get an objective view of the project threats because different projects have different risk thresholds and different risk appetite. In this article the fuzzy approach to risk appetite is suggested which allow to incorporate fuzziness of this evaluation and multiple criteria approach based on the different preferences of the project board or owners of the project and different types of the risk appetite.

Keywords: Risk appetite, task criticalness potential, threatening task, multiple criteria decision making.

JEL Classification: O22, C44
AMS Classification: 90B50, 90B99

1 Introduction

The project management is a tool providing successful realization of a project within planned time, planned costs and planned quality. To ensure this it is necessary to analyze critical and threatening factors both explicitly and implicitly existing in the project.

The riskiness, threat and criticalness of project tasks are not only given by the surroundings and the environment of the project but also by the internal arrangement and structure of the project.

Quantitative estimation serves for the evaluation of the task criticalness potential. Although such quantitative estimation of the project is performed in a pre-project phase, generally only traditional time and resource analysis of the project is provided. Primarily, the tasks are evaluated from a time perspective only as critical or uncritical tasks, but many other quantitative characteristics can be used. Such a multiple attribute analysis enables not to ignore the impact of time-uncritical tasks, which may have the high criticalness from many different reasons [3].

The project risk deals with risk management, however, without a quantitative view of the sequence and arrangement of project tasks and its other quantitative qualities. In the area of risk management, a number of different techniques and approaches have been derived in order to reduce the risk of a project and its partial tasks [13]. The criticality of project tasks is often defined from a time perspective only, using stochastic approaches [2] and [8], fuzzy sets methods [6], [15] or using the findings of a network analysis [9] and [4]. Gong and Rowings [9] mention that ignoring the impact of non-critical tasks, which may easily become critical, is the most frequent criticism of project duration analysis methods. Another point of view on tasks criticalness is given by the structure of relations in the project. Bowers [2] or Williams [14] deal with a stochastic analysis of a project network where the criticality of tasks in the project is derived from the relation between task duration and the whole project, and on the basis of a number of resources used for a task and the whole project.

The main aim of the paper is to compare the qualitative and quantitative approach analyzing threats and criticalness potential of the project tasks. We suggest the application of the fuzzy linguistic system into the qualitative (soft) estimation of the threat of the project tasks and we use multiple attribute approaches in the quantitative (hard) evaluation of the tasks criticalness potential. We compare both approaches and discuss their advantages and disadvantages.

[1] Czech University of Life Sciences, Faculty of Economics and Management, Dept. of Systems Engineering 165 21 Praha 6 – Suchdol, brozova@pef.czu.cz
[2] Czech University of Life Sciences, Faculty of Economics and Management, Dept. of Systems Engineering 165 21 Praha 6 – Suchdol, bartoska@pef.czu.cz
[3] Czech University of Life Sciences, Faculty of Economics and Management, Dept. of Systems Engineering 165 21 Praha 6 – Suchdol, subrt@pef.czu.cz

2 Method for classification of the task threat

Suggested method for the task classification is based on the evaluation of the so-called task criticalness potential and its fuzzyfication using a proper fuzzy linguistic scale.

2.1 Evaluation of the task criticalness potential

A hard system approach to the analysis of the task in the relation to the project success is based on the crisp exact quantitative evaluation of the task. Brožová et al [3] suggested providing the overall evaluation of the task criticalness potential without soft knowledge about character of the tasks.

The estimation of the overall criticalness potential of the project tasks is based on the multiple attributes decision-making method using 5 indicators of the criticalness as topological location, which reflects the probability the critical path will pass through the current task, time duration, time slack, cost, and work. The last four indicators are evaluated by project schedule. The weights of these indicators are evaluated by the decision maker.

The values of these indicators are then converted into interval $\langle 0,1 \rangle$, the value 0 corresponds to the lowest task criticalness and the value 1 to the highest criticalness, and used as the input for the Simple additive weighted method [10] using formula

$$C_i = v_1 \frac{p_i - \min\limits_K p_K}{\max\limits_K p_K - \min\limits_K p_K} + v_2 \frac{t_i - \min\limits_K t_K}{\max\limits_K t_K - \min\limits_K t_K} + v_3 \frac{s_i - \max\limits_K s_K}{\min\limits_K s_K - \max\limits_K s_K} + v_4 \frac{c_i - \min\limits_K c_K}{\max\limits_K c_K - \min\limits_K c_K} + v_5 \frac{w_i - \min\limits_K w_K}{\max\limits_K w_K - \min\limits_K w_K} \qquad (1)$$

where C_i is the criticalness potential of the task i, p_i is the topological location which reflects the probability the critical path will pass through the current task i, t_i is the duration of the task i, s_i is the time slack of the task i, c_i is the cost of the task i, v_i is the work amount of the task i, and $v_1,\ldots,v_5$ are the weights of the components of the criticalness potential.

The topological location indicator, which reflects the probability the critical path will pass through the current task i, is related to the project structure, these indicators of the predecessors j and the numbers of successors of tasks j, are calculated as

$$p_0 = 1 \quad \text{and} \quad p_i = \sum_{j\ predecessors\ i} \frac{p_j}{h_j},\ i = 1,2,...,N \qquad (2)$$

where p_0 is the probability of starting a dummy task is a component of the critical path, p_0, p_i, p_j are the probability of a critical path gateway through tasks 0, i, i, and h_j is the number of tasks following task j.

2.2 Fuzzy approach to project threats

The suggested classification of the threat of the project tasks is based on experts' knowledge and experts' experience of characters of tasks and project using the soft system approach [5] because many factors influencing the project success have social or human roots or are based on uncertainty, probability and possibility. Such managerial evaluation of the task threat can be made by expressing linguistic uncertainty using linguistic terms and fuzzy scales ([11], [12]). A fuzzy measure is a subjective scale for the degrees of fuzziness and is suitable for analyzing human subjective judgment [7]. Therefore, we define the fuzzy linguistic system with the linguistic state variables describing the level of task threat, their values are words, and meanings of these words are fuzzy sets. In this study, we created three uniform and non-uniform five point fuzzy scale with terms T1, T2, T3, T4 and T5.

Uniform scale is used in the case of a risk-neutral appetite of the project board or owners of the projects.

Task can be	
T1 - Non-threatening	(0; 0; 0.11; 0.22)
T2 - Weakly threatening	(0.11; 0.22; 0.33; 0.44)
T3 -Rather threatening	(0.33; 0.44; 0.56; 0.67)
T4 -Strongly threatening	(0.56; 0.67; 0.78; 0.89)
T5 - Extremely threatening	(0.78; 0.89; 1; 1)

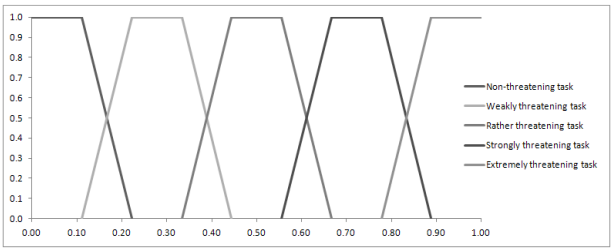


Figure 1 Linguistic uniform fuzzy scale for overall task threat evaluation

Non-uniform scale with uncertainty increasing towards the evaluation Extremely threatening task is suitable for a risk-averse appetite of the project board or owners of the project.

Task can be	
T1 - Non-threatening	(0; 0; 0.0; 0.1)
T2 - Weakly threatening	(0.0; 0.1; 0.2; 0.3)
T3 -Rather threatening	(0.2; 0.3; 0.45; 0.55)
T4 -Strongly threatening	(0.45; 0.55; 0.7; 0.8)
T5 - Extremely threatening	(0.8; 0.9; 1; 1)

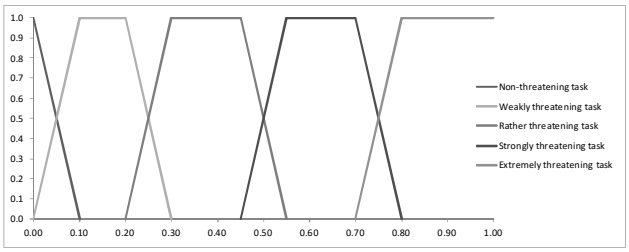


Figure 2 Linguistic fuzzy scale for overall task threat evaluation for the risk-averse appetite

Non-uniform scale with uncertainty increasing towards the evaluation Non-threatening task is suitable for the risk-prone appetite (risk tolerance) of the project board or owners of the project.

Task can be	
T1 - Non-threatening	(0; 0; 0.2; 0.3)
T2 - Weakly threatening	(0.2; 0.3; 0.45; 0.55)
T3 -Rather threatening	(0.45; 0.55; 0.7; 0.8)
T4 -Strongly threatening	(0.7; 0.8; 0.9; 0.1)
T5 - Extremely threatening	(0.9; 0.1; 1; 1)

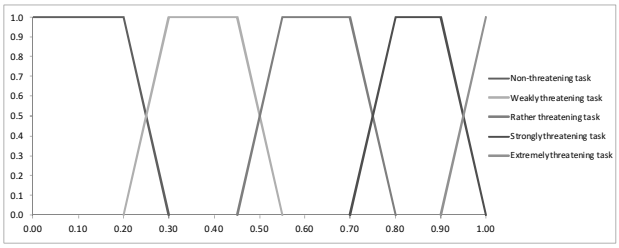


Figure 3 Linguistic fuzzy scale for overall task threat evaluation for the risk-prone appetite

The suggested fuzzy scales are then used for classification of the project task according to their criticalness potential into the groups of the task with the same fuzzy task threat.

The task criticalness potential provides crisp measurements, and for suggested task classification, these crisp values must be transformed into linguistic terms in a selected fuzzy scale, which reflects properly the risk appetite of the project managers. The measure of the task threat is evaluated by a linguistic term *Tj*, if the degree of membership of the value of the task criticalness to given term is equal to 1.

$$\text{IF } \mu_{Tj}(C_i) = 1 \text{ THAN Task } i \text{ is } Tj, \text{ for all } i = 1,2,...,N, \quad j = 1,2,...,5 \tag{3}$$

In other case, the task threat is evaluated as between *Tj* and *Tj*+1.

$$\begin{aligned} &\text{IF } \mu_{Tj}(C_i) > 0 \text{ AND } \mu_{Tj+1}(C_i) > 0 \\ &\text{THAN Task } i \text{ is between } Tj \text{ and } Tj+1, \text{ for all } i = 1,2,...,N, \quad j = 1,2,...,4 \end{aligned} \tag{4}$$

where C_i is the criticalness potential of the task i, $\mu_{Tj}(C_i)$ is the membership function of the fuzzy set corresponding to the term *Tj*, *N* is a number of the tasks.

3 Example of the evaluation of the tasks

To evaluate a project as a whole from the point of view of its task criticalness is not easy. An unambiguous and sufficient approach is still non-existent. Tasks on a critical path are considered the most threatening tasks. The finding of a critical path and a follow-up analysis of reserves does not always sufficient. In practice, it can often be seen that even tasks off a critical path have an extreme impact on the realization and success of a project. The following small-scale project is used as an illustrative example of the suggested method for task classification. This example illustrates the approach of the assessment of the task threat but nevertheless its small-scale shows the benefits that the use of fuzzy tools has.

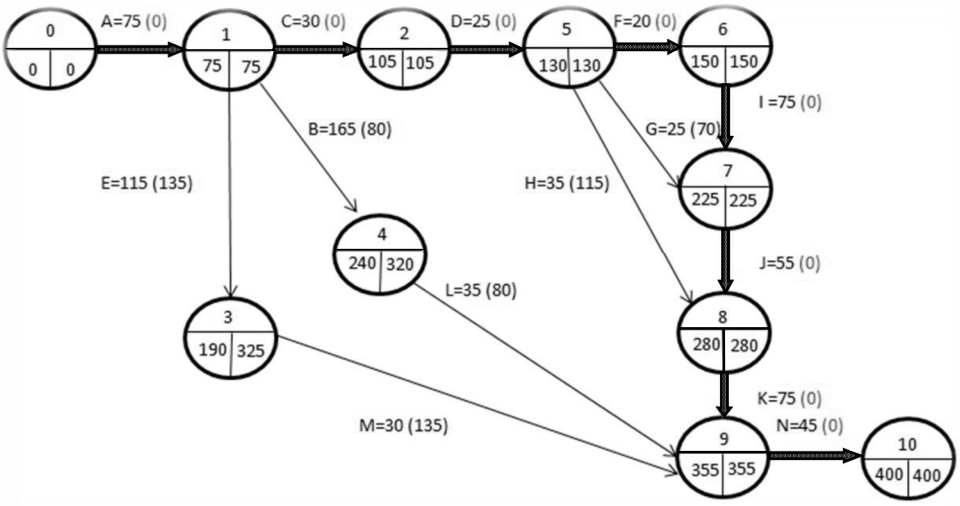


Figure 4 Small-scale projects with the indication of a critical path

The project was firstly evaluated using quantitative data. Table 1 presents partial characteristics of the task criticalness and criticalness potential towards a project goal (using formula 1 and 2). The illustrative example was used from authors' previous research [3]. The weights used for the computation of the criticalness potential were taken from authors' previous research also [1].

Task	Time criticalness	Topological criticalness	Slack criticalness	Cost criticalness	Work criticalness	Criticalness potential
A	0.379	1	1	0.656	0.198	0.615
B	1	0.247	0.407	0.473	0.221	0.451
C	0.069	0.247	1	0.290	0.290	0.337
D	0.034	0.247	1	0.059	0.059	0.212
E	0.655	0.247	0	0.321	0.145	0.280
F	0	0	1	0	0	0.129
G	0.034	0	0.481	0.046	0.046	0.092
H	0.103	0	0.148	0.076	0.076	0.076
I	0.379	0	1	0.771	0.771	0.590
J	0.241	0.124	1	0.557	0.557	0.481
K	0.379	0.247	1	1	1	0.756
L	0.103	0.247	0.407	0.415	0.415	0.331
M	0.069	0.247	0	0.153	0.061	0.116
N	0.172	1	1	0.450	0.450	0.580
Weights	0.164	0.189	0.129	0.288	0.230	

Table 1 Analysis of the task criticalness using Simple additive weighted method

4 Results and discussion

Tasks are now classified according to its criticalness potential. Firstly a neutral fuzzy linguistic scale for a risk appetite was applied (Figure 1).

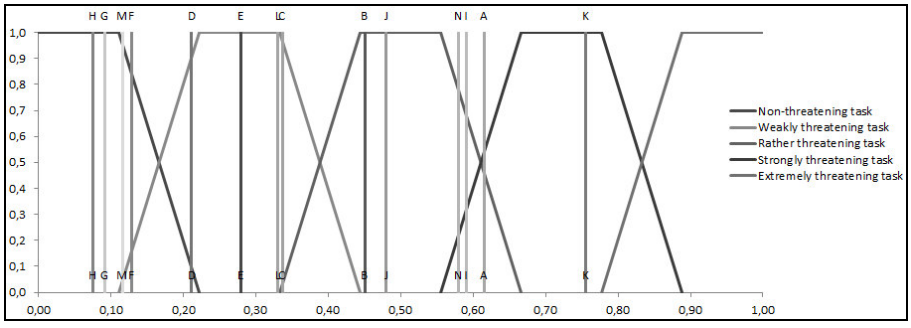


Figure 5 The risk-neutral appetite of the project board

If a project owner or project board is not diverted to either side in the risk appetite, or is not impartial, i.e. is averse or prone to the threats, the project manager can express the task threats towards the risk appetite using neutral fuzzy scale. Figure 5 shows small-scale project tasks expressed using the linguistic uniform fuzzy scale for the risk-neutral appetite. The project manager will regard task K as Strongly threatening for the project, while activities N, I and A will be regarded between Strongly and Rather threatening tasks and so on.

If, within the project's assignment, an emphasis is put on a low level of threat and if the project owner or project board are not willing to take risks and undergo threats, the project manager could use the linguistic fuzzy scale for the risk-averse appetite (Figure 2). This is demonstrated in Figure 6 for the illustrative example. A number of activities, which can be designated on this scale as Extremely or Strongly threatening, are growing (Table 2). The project manager would be constrained to plan and carry out preventive measures, or prepare crisis scenarios. Activities N, I, A are Strongly threatening tasks and K is in this case considered between Strongly and Extremely threatening. Expenses for risk management and defence against risks in the project would grow.

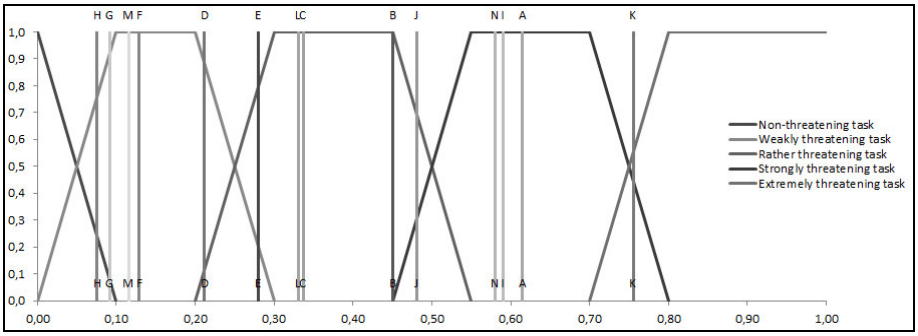


Figure 6 The risk-averse appetite of the project board

If the project manager or project board assigning the project are willing to undergo a higher level of threat and are rather unafraid of threats, a number of activities in the illustrative example which can be marked as Extremely or Strongly threatening is falling (see Figure 7). The project manager would not be constrained to put and increased effort into planning defence, preventive measures or preparing crisis scenarios. Activities N, I, and A are only Rather threatening tasks and task K (see Table 2) can be regarded between Rather and Strongly threatening. In contrary tasks H, G, M and F are seen as Non-threatening tasks.

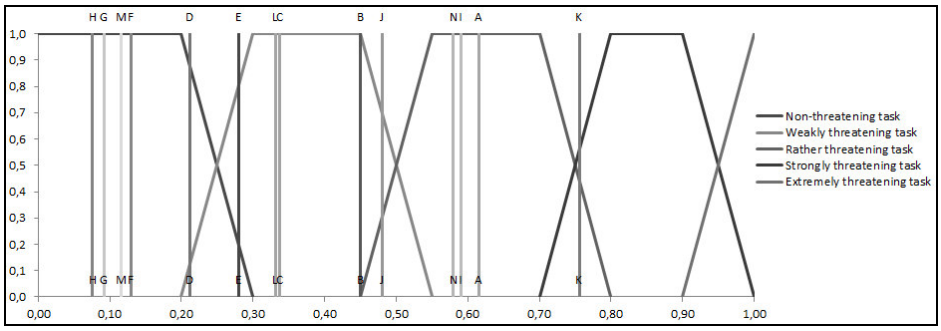


Figure 7 The risk-prone appetite of the project board

Table 2 compares results of the three above-mentioned approaches, i.e. the risk-neutral appetite (Figure 5), risk-averse appetite (Figure 6) and risk-prone appetite (Figure 7). In the case of the risk-averse appetite there is an increase in a number of tasks designated as Strongly or Extremely threatening while in the case of the risk-prone appetite (risk tolerance) a number of such tasks is falling.

The calculated value of criticalness potential for individual tasks in the illustrative example shows tasks placement on the linguistics fuzzy scale (Table 2). The order of tasks in terms of treat is identical; however, there is a shift of the tasks on their classification. Using the proper linguistic fuzzy scale, the criticalness potential value is interpreted towards the risk appetite of the project owner or project board.

	Risk-prone appetite	Risk-neutral appetite	Risk-averse appetite
T1 - Non-threatening task	H G M F	H G	
Between T1 and T2	D E	M F D	H G
T2 - Weakly threatening task	L C B	E L C	M F
Between T2 and T3	J		D E
T3 - Rather threatening task	N I A	B J	L C B
Between T3 and T4	K	N I A	J
T4 - Strongly threatening task		K	N I A
Between T4 and T5			K
T5 - Extremely threatening task			

Task	Criticalness potential	Task	Criticalness potential
H	0.076	J	0.481
G	0.092	N	0.580
M	0.116	I	0.590
F	0.129	A	0.615
D	0.212	K	0.756
E	0.280		
L	0.331		
C	0.337		
B	0.451		

Table 2 Final classification of project tasks based on three linguistic fuzzy scales

5 Conclusion

The paper brings an original approach to the expression of risk appetite for a project owner or project board. First, in this approach the criticalness potential for each task of a project is calculated. The criticalness potential is then interpreted on three different linguistic fuzzy scales for three ways of expressing risk appetite describing an expectation of the project board. The paper suggest different linguistic fuzzy scales, which are used on an illustrative example for risk-neutral appetite, risk-averse appetite and risk-prone appetite (risk tolerance) respectively. The paper offers a particular suggestion for the practice of managing project risks (risk project management). Applying the linguistic fuzzy scale, project tasks can be distinguished in terms of their threat in a direct context of the risk appetite.

Acknowledgements

The research is supported by the Internal Grant Agency of the University of Life Sciences Prague – project IGA PEF 20131028 - Multicriterial Analysis of Criticality of Project Activities.

References

[1] Bartoška, J., Brožová, H., Šubrt, T. and Rydval, J.: Incorporating practitioners' expectations to project management teaching. In: *Efficiency and responsibility in education 2013*. Czech University of Life Sciences Prague, Prague, 2013, 16–23.

[2] Bowers, J.: Identifying Critical Activities in Stochastic Resource Constrained Networks. *Omega - International Journal of Management Science* **24** (1996), 37–46.

[3] Brožová, H., Bartoška, J., Šubrt, T. and Rydval, J.: Multiple criteria evaluation of the activities criticalness in the Project management. In: *Proceedings of the 31st International conference on Mathematical Methods in Economics*. College of Polytechnics Jihlava, Jihlava, 2013, 83–88.

[4] Chanas, S. and Zielinski, P.: On the hardness of evaluating criticality of activities in a planar network with duration intervals. *Operation Research Letters* **31** (2003), 53–59.

[5] Checkland, P. and Scholes, J.: *Soft System Methodology in Action*. John Wiley & Sons Inc., Chichester, 1990.

[6] Chen, C. T. and Huang, S. F.: Applying fuzzy method for measuring criticality in project network. *Information Sciences* **177** (2007), 2448–2458.

[7] Chen, T-Y.: Using a Distance between Fuzzy Number-Valued Fuzzy Measures to Determine Attribute Importance: An Outcome-Oriented Perspective. *International Journal of Fuzzy Systems* **8** (2006), 127-136.

[8] Cruz, S., García, J. and Herrerías, R.: Stochastic models alternative to the classical PERT for the treatment of the risk: mesokurtic and of constant variance. *Central European Journal of Operations Research* **7** (1999), 159–175.

[9] Gong, D. and Rowings, J. E.: Calculation of safe float use in risk-analysis-oriented network scheduling. *International Journal of Project Management* **13** (1995), 187–194.

[10] Hwang, Ch. L. and Yoon, K.: *Multiple Attribute Decision Making*. Springer Verlang Berlin Heidelberg, New York, 1981.

[11] Ross, T. J.: *Fuzzy Logic with Engineering Applications*. John Wiley & Sons Inc., Chichester, 2010.

[12] Wang, Z. and Klir, G. J.: *Fuzzy Measure Theory*. Plenum Press, New York, 1992.

[13] Wideman, R. M.: *Project and Program Risk Management*. Project Management Institute, 1992.

[14] Williams, T. M.: Criticality in stochastic networks. *Journal of Operational Research Society* **43** (1992), 353–357.

[15] Yakhchali, S. H.: A path enumeration approach for the analysis of critical activities in fuzzy networks. *Information Sciences* **204** (2012), 23–35.

A Comparison of Bayesian Methods for Estimating of Multivariate Models with Censored Data

Jan Brůha[1]

Abstract.

Recently, multivariate models with censored data have been proposed in econometrics with a large set of applications that range from financial econometrics to the analysis of individual behaviour. Typically, these models are estimated either using methods of moments or by Gibbs samplers in the Bayesian framework.

In this paper, I discuss Bayesian alternatives based on the importance sampling. The importance sampling approach has the advantage of generating independent samples, hence avoiding the problems with autocorrelation of Gibbs samplers. On the other hand, importance sampling requires likelihood evaluation which is difficult for multivariate models with censored data. I therefore also propose the extension of the parameter space and the sampling density. Such an extension is helpful in avoiding the likelihood evaluation. Finally, I compare the standard Gibbs sampler with the proposed importance sampling algorithm for speed and accuracy.

Keywords: Multivariate Tobit model, Gibbs sampler, Importance Sampling

JEL classification: C11; C35
AMS classification: 62F15

1 Introduction

Recently, multivariate models with censored data have been proposed in econometrics with a large set of applications that range from financial econometrics to the analysis of individual behaviour. These models can be considered as extensions of the original Tobin work ([13]) to the multivariate setting ([9]), and further to the multivariate setting with non-gaussian errors ([12]) or mixture models ([1]).

These models have a natural latent variable representation. Given latent variables, the likelihood function is easily computed, but without them, the likelihood function leads to the evaluation of integrals that typically are not available in closed forms. Therefore, especially in high dimensions, the estimation of such models is done using methods of moments ([5]). In the Bayesian framework, the difficulty of likelihood evaluation means that even unnormalised posterior is difficult to evaluate too. Hence, importance sampling or Metropolis-Hastings algorithms seem impractical, and Gibbs samplers have been proposed ([1], [9], [10], [12]). Moreover, the latent-data structure of the model makes the Gibbs sampler a natural option.

In this paper, I propose a Bayesian alternative based on the importance sampling. The importance sampling approach has the advantage of generating independent samples, hence avoiding the problems with autocorrelation of Gibbs or Metropolis-Hastings samplers. Nevertheless, importance sampling requires the likelihood evaluation which -as said- is difficult for multivariate models with censored data. I therefore propose the following modification: not only parameters, but also latent variables are sampled and the careful choice of both sampling densities is crucial for the competitiveness of the proposed algorithm. I compare the usual Gibbs sampler with the proposed importance sampling algorithm for speed and accuracy.

[1]Kolin Institute of Technology, Okružní 703, Kolín, Czech Republic, jan_ bruha@yahoo.co.uk

2 Model formulation

The original Tobit model is a statistical model proposed by [13] to describe the relationship between a non-negative dependent variable y and a vector of covariates X. The Tobit model is cast in the latent (unobserved) variable framework. It is assumed that there is a latent variable y^*, which depends on X via the linear regression function ($y^* = X\beta + u$), and the observed variable is defined to be equal $y = \max(y^*, 0)$, i.e., y is equal to y^* if the latent variable is positive and $y = 0$ otherwise. The Tobit model has been applied in statistics to modelling non-negative observations, such as infrequent purchases (of durable goods or luxury goods). The goal of the estimation is to estimate the parameter vector β, which are useful for making predictions about quantities of interest[1] for an individual i with observed characteristics X_i.

Recently, multivariate extensions of the Tobit model have been proposed. Let Y_i be a D-dimensional vector of non-negative numbers, and let X_i be covariates of individual i. Then, the model can be formulated as:

$$Y_i = \max(Y_i^*, \mathbf{0}), \tag{1}$$

where

$$Y_i^* = X_i\beta + u_i, \tag{2}$$

$\mathbf{0}$ is the D-dimensional vector of zeros, β is the vector or matrix of (unknown) parameters, and u_i is a D dimensional error terms with a zero mean and the covariance matrix $\Sigma = (\sigma_{ij})$. The operator max in (1) is applied component-wise.

In the rest of the paper, I assume that u has a multivariate normal (henceforth MVN) distribution (this assumption is relaxed by [12] and the algorithm presented in Section 3.2 can be easily adapted for such an extension): $u \sim N(\mathbf{0}, \Sigma)$. Then Y^* is also a random vector with the MVN distribution with the mean $X_i\beta$ and the covariance matrix Σ.

3 Estimation approaches

One direct attempt to estimate the model (1) - (2) is to maximize the likelihood function. Consider therefore the likelihood function for the observation j and denote ℓ_j those elements of Y_j which are strictly positive (and the complement $\widetilde{\ell}_j$ which are zero). The contribution of the observation j to the likelihood can be written as:

$$\mathbb{L}_j = \text{Prob}(Y_{\widetilde{\ell}_j j} = 0 | Y_{\ell_j j}) \times \phi(Y_{\ell_j j} | \mu_{\ell_j j}, \Sigma_{\ell_j})$$

where $\phi(*|m, S)$ is the density function of the MVN distribution with the mean m and the covariance matrix S, $\mu_{\ell_j j}$ is the vector which includes entries corresponding to strictly positive observations ℓ_j from the vector $\mu_j = X_j\beta$, Σ_{ℓ_j} is the analogous selection from the matrix Σ (only rows and columns corresponding to strictly positive observations ℓ_j are included), and $\text{Prob}(Y_{\widetilde{\ell}_j j} = 0 | Y_{\ell_j j})$ is the conditional probability that we observe zeros for elements $\widetilde{\ell}_j$ given the positive observations $Y_{\ell_j j}$.

The conditional probability $\text{Prob}(Y_{\widetilde{\ell}_j j} = 0 | Y_{\ell_j j})$ is the probability that a MVN random vector with parameters $\widetilde{\mu}_j$ and $\widetilde{\Sigma}_j$ will have all the entries non-positive.[2] The evaluation of the probability that a MVN vector will have all entries non-positive is difficult. From the technical point of view, this probability is the multivariate integral over a rectangle and this is not an easy task unless $\widetilde{\Sigma}$ is a diagonal matrix[3]. Although there has been a huge progress in recent years in the evaluation of such integrals, this is still a computationally intensive task([7]). This means that likelihood-based methods are probably not practical in high dimensions and if there is a significant number of observations with many zeros.

Moreover, also Bayesian methods that requires the evaluation of the (unnormalised) posterior densities (i.e. the product of the prior density and the likelihood) seem impractical too. Hence importance

[1] These quantities of interest include the probability that the positive observation will occur (probability that $y_i > 0$), the expected value of y_i, or the conditional expectation $\mathbb{E}(y_i | y_i > 0)$ and so on.

[2] The parameters $\widetilde{\mu}$ and $\widetilde{\Sigma}$ are parameters of the conditional MVN distribution (conditional on the observed values $Y_{\ell_j j}$) and are given as follows: $\widetilde{\mu}_j = \mu_{\widetilde{\ell}_j j} + \Sigma_{\widetilde{\ell}_j \ell_j} \Sigma_{\ell_j \ell_j}^{-1} (Y_{\ell_j j} - \mu_{\ell_j j})$ and $\widetilde{\Sigma} = \Sigma_{\widetilde{\ell}_j \widetilde{\ell}_j} - \Sigma_{\widetilde{\ell}_j \ell_j} \Sigma_{\ell_j \ell_j}^{-1} \Sigma_{\ell_j \widetilde{\ell}_j}$.

[3] In that case, the integral is given as the product of the commutative density functions of the univariate normal distributions.

samplers and Metropolis-Hastings algorithms are typically avoided. On the other hand, the latent-data representation makes the Gibbs sampler a natural option that can be easily coded. Nevertheless, in this paper, I propose an importance sampling algorithm that can be used to estimation of the posterior density of such models.

3.1 Gibbs sampler

Given the latent-data structure of the model, the natural Bayesian choice is the Gibbs sampler. The idea is simple: if the latent variables Y^* are observed, then the Bayesian estimation of the parameters β and Σ is basically reduced to the estimation of the seemingly unrelated regression (SUR) model, and there are efficient algorithms for Bayesian estimation of the SUR model. However, if the parameters β and Σ are known, then it is possible to sample Y^*.

The immediate idea would be as follows. Iterate the sampler between (i) sampling of the parameters β and Σ using draws of Y^* as data; and (ii) sampling from the conditional distribution of Y^* given actual observations Y and the current sample of the parameters β and Σ. The full sampling from the distribution of Y^* given actual observations Y and parameters would be time-consuming[4]. To reduce the computational time, one needs to realise that it is not necessary to sample from the conditional distribution of Y^*, but it is sufficient to sample from Y^* sequentially (see Algorithm 2 below).

I provide details for the sampling from the posterior distribution for the independent normal Wishart prior, i.e., the prior on β is specified as a MVN distribution, while the prior on Σ is an inverted Wishart distribution. In what follows, I denotes the sample size (the number of units), while D denotes the dimensionality of the model (i.e., the number of components of Y_i).

Algorithm 1

- Set priors for β and Σ; the prior for β is MVN with parameters β_0 and S_0, the prior for Σ is an inverted Wishart distribution with parameters c_0 and C_0^{-1}. Set the burn-in periods N_1 and genuine iterations N_2.
- Set initial values of the parameters $\beta^{(1)}$ and $\Sigma^{(1)}$ (for example by drawing from the prior distributions if they are proper).
- Set $n = 1$ and run the following Gibbs sampler:
  1. Given $\beta^{(n)}$ and $\Sigma^{(n)}$ and Y, sample $Y^{*(n)}$ using Algorithm 2 described below.
  2. Given the current sample of $Y^{*(n)}$ and $\Sigma^{(n)}$, sample $\beta^{(n+1)}$ from MVN distribution with the mean $\widehat{m}$ and the covariance matrix $\widehat{B}$, where $\widehat{m} = \widehat{B}\left[\sum_{i=1}^{I} X_i'(\Sigma^{(n)})^{-1}Y_i^{*(n)} + S_0^{-1}\beta_0\right]$ and $\widehat{B} = \left[\sum_{i=1}^{I} X_i'(\Sigma^{(n)})^{-1}X_i + S_0^{-1}\right]^{-1}$;
  3. Given the current sample of $Y^{*(n)}$ and $\beta^{(n+1)}$, sample $\Sigma^{(n+1)}$ from the inverted Wishart distribution with the following parameters: $c_1 = c_o + I$ and $C_1^{-1} = C_0^{-1} + \sum_{i=1}^{I}(Y_i^{*(n)} - X_i\beta^{(n+1)})(Y_i^{*(n)} - X_i\beta^{(n+1)})'$;
  4. Set $n = n + 1$ and repeat until $n = N_1 + N_2$.
- Disregard the first N_1 draws and use the remaining N_2 samples as the sample from the posterior distribution for β and Σ.

Steps (2) and (3) are standard steps of the Bayesian estimation of seemingly unrelated regressions ([6]).

Algorithm 2

- Given the parameters $\beta^{(n)}$ and $\Sigma^{(n)}$ and data Y; sample latent data Y^* as follows:

[4]It is not an easy task to sample from the truncated multivariate normal distribution. There are various possibilities of doing this: for example the obvious accept-reject algorithm, which is time-consuming if the probability of positive observation is not high. The alternative is another Gibbs sampler, which would iterate over conditional marginal densities of Y^*. Under this possibility, the main Gibbs sampler would incorporate another sampler.

1. For $i = 1, \dots I$ do the following
2. For $d = 1, \dots D$ do the following
3. If $Y_{id} > 0$ set $Y_{id}^{*(n)} = Y_{id}$, else derive the conditional distribution of $Y_{id}^{*(n)}$ which is the normal distribution truncated at zero from the left with the parameters[5] $\widetilde{\mu}_{id}$ and $\widetilde{\sigma}_{id}^2$. The sample from the truncated distribution is simple, e.g., the inverse method gives $\widetilde{\mu}_{id} + \widetilde{\sigma}_{id}\Phi^{-1}(u * \Phi(-\widetilde{\mu}_{id}/\widetilde{\sigma}_{id}))$, where Φ is the cumulative distributive function of the *standard* normal distribution, Φ^{-1} is the quantile function of the *standard* normal distribution, and u is a draw from the uniform random variable on the interval [0, 1].

Note that this algorithm **is not** an algorithm to sample from the joint density of Y^* given the current samples of the parameters β and Σ. In fact, the sample values of say Y_{id}^* are conditional on the current samples of the parameters (β, Σ), on the current sample of $Y_{i\delta}^*$ for $\delta < d$, and on the past samples of $Y_{i\widetilde{\delta}}^*$ for $\widetilde{\delta} < d$. However, due to the construction of the Gibbs sampler, this is sufficient.

Some authors suggest in a related context ([11]) to sample the latent variables from the joint conditional distribution. This can be (approximatively) achieved by M-times repetition ($M > 1$) of Algorithm 2 in Step 1 of the Gibbs sampler. My Monte Carlo experiments suggest that this approach does not reduce the autocorrelation of the sampled parameters β and Σ and does not lead to savings of computational times or accuracy gains. Therefore, I do not pursue this option in the rest of the paper.[6]

3.2 A proposed importance sampler

The importance sampling approximation to the posterior distribution is based on sampling from an auxiliary distribution and the sample is corrected by the weights given by the Radon-Nikodym derivative of the posterior distribution with respect to the auxiliary distribution. It is well known that if $\pi_{\text{post}}(\theta)$ is the posterior distribution of interest, and $\wp(\theta)$ is the auxiliary distribution, the posterior expectation of any integrable function[7] f can be under weak conditions ([6]) approximated as :

$$\mathbb{E}_{\text{post}} f(\theta) \cong \frac{1}{\sum_{n=1}^{N} \omega^{(n)}} \sum_{n=1}^{N} \omega^{(n)} f(\theta^{(n)}), \tag{3}$$

where $\theta^{(n)}$ are samples from the auxiliary distribution $\wp$ and the weights $\omega^{(n)}$ are given as:

$$\omega^{(n)} = \frac{\tilde{\pi}_{\text{post}}(\theta^{(n)})}{\wp(\theta^{(n)})}, \tag{4}$$

where $\tilde{\pi}_{\text{post}}$ is a kernel (i.e. unnormalized version) of the posterior distribution π_{post}.

The quality of approximation of (3) depends on how $\wp$ is close to the distribution of interest π_{post}. In fact, the importance sampling may fail completely if the auxiliary distribution does not cover important parameter space for the posterior. This is one of the difficulty with importance sampling – finding an appropriate auxiliary distribution, which can be easily sampled but which is close to the posterior distribution.

In the context of multivariate models with censored data, there is another difficulty. Even the kernel of the posterior is given as the product of the kernel of the prior distribution and the likelihood, but the likelihood is not easy to evaluate in these models. Hence, the computation of weights in (4) is time consuming.

[5] The parameters of the conditional distribution are given as follows: $\widetilde{\mu}_{id} = \mu_{id}^{(n)} + \Sigma_{d,-d}^{(n)}(\Sigma_{-d,-d}^{(n)})^{-1}(Y_i^{*(c)} - \mu_{i,-d}^{(n)})$, $\widetilde{\sigma}_{id}^2 = \Sigma_{d,-d}^{(n)} - \Sigma_{d,-d}^{(n)}(\Sigma_{-d,-d}^{(n)})^{-1}\Sigma_{-d,d}^{(n)}$, where $\mu_{id}^{(n)}$ is the dth element of the unconditional mean $X_i\beta^{(n)}$, $\Sigma_{d,-d}^{(n)}$ is the dth row of the unconditional covariance matrix $\Sigma^{(n)}$, $\Sigma_{-d,-d}^{(n)}$ is the submatrix of $\Sigma^{(n)}$ after deleting the dth row and dth column, and $\mu_{i,-d}^{(n)}$ is created from the unconditional mean $X_i\beta^{(n)}$ by deleting the dth row. Finally, $Y_i^{*(c)}$ contains the current sample from Y_i^* and its elements are given as $Y_{i\delta}^{*(c)} = Y_{i\delta}^{*(n+1)}$ for $\delta < d$ and $Y_{i\delta}^{*(c)} = Y_{i\delta}^{*(n)}$ for $\delta > d$.

[6] Details available upon request.

[7] By suitable choices of f, one can evaluate almost any quantity of Bayesian interests, such as the posterior mean, the posterior variance, quantiles of the posterior distributions etc.

The idea of this paper is to use the latent-data structure of the model and to consider *two* auxiliary sampling densities: (i) one is the 'standard' density for the true parameters of the model, and (ii) second one is for latent variables.

Hence, I propose the following algorithm: ALGORITHM 3

1. Set the size of the sample N and proposal densities. For $n = 1, ..., N$, do the following:
2. Sample θ^n from the auxiliary density $\hbar$
3. Sample $Y^{*(n)}$ from another auxiliary density $\upsilon(*|Y, \theta^n)$
4. Evaluate sample weights as follows:

$$\omega^{(n)} = \frac{\pi_{\text{pri}}(\theta^{(n)})\mathbb{L}(\theta^{(n)}|Y^{*(n)})}{\hbar(\theta^{(n)})\upsilon(Y^{*(n)}|Y, \theta^n)}.$$

The idea is to replace the need of evaluation of difficult task $\mathbb{L}(\theta^{(n)}|Y)$ by much simpler task $\mathbb{L}(\theta^{(n)}|Y^{*(n)})$. Indeed, while the evaluation of $\mathbb{L}(\theta^{(n)}|Y)$ requires a multidimensional integration over a nasty region, the evaluation of $\mathbb{L}(\theta^{(n)}|Y^{*(n)})$ is almost trivial. In the context of this paper, it is just an evaluation of MVN probability density function.

The crucial choice is the choice of importance densities $\hbar$ and $\upsilon(Y^*|Y, \theta^n)$. The choice of $\hbar$ is relatively standard choice and in general depends on the context. The actual crucial choice is the choice of the density $\upsilon(Y^*|Y, \theta^n)$. In this paper, I propose to based it on the product of marginal components of the true conditional density $p(Y^*|Y, \theta^n)$. This has the advantage of the covering the whole state space, but on the other hand, the simulation from such a density is simple: simulation from the product of marginal densities can be obviously achieved by simulating individual components separately and simulation from the *univariate* truncated density is simple – it can be easily coded using the inverse technique.

3.3 Comparison

In this part of the paper, I comment results of a Monte Carlo experiments that can be used to compare the relative merits of the two approaches. In fact, I found that the importance-sample approach is practical only for relatively small datasets (up to 100 observations). Otherwise, the Gibbs sampler approach is better in computational time (to achieve a given level of accuracy). Hence, I conclude that the alternative approach is a useful way forward of thinking about Bayesian computation, but subject to future research. The reason why the alternative does not prove much better is the relatively low autocorrelation of the Gibbs sampler in this particular application. This means that in other applications with latent variables it is still worth exploring the potential of formulating the estimation problems using independence samplers.

4 Conclusions

In this paper, I propose a Bayesian algorithm for estimation of multivariate models with censored data. The proposed algorithm is based on importance sampling and it is an alternative to the usual Gibbs sampler. I show that for small or middle samples, the proposed algorithm is competitive. Moreover, the importance sampling structure can benefit from advances in research on the importance sampling approach, such as adaptive proposal density modification ([2, 4]) or sequential Bayesian estimation ([3]).

The proposed algorithm makes use of the latent-data formulation of multivariate models with censored data. Nevertheless, the algorithm formulation is general enough to be useful for other econometric models which has a natural latent-data representation. This direction can be pursued in future research.

Acknowledgements

The research on this paper has been supported by the grant of the Czech Scientific Foundation no. 14-22932S. The support is gratefully acknowledged.

References

[1] Brůha J., and Brůhová-Foltýnová, H.: Active Travel in Prague: Assessing Heterogeneity in the Population. Working paper, Kolin Intitute of Technology, available at http://www.koltech.cz/english/data/Active-travel-Prague.pdf, 2013.

[2] Cappé, O., Guillin, A., Marin, J., and Robert, C.: Population Monte Carlo. *Journal of Computational and Graphical Statistics* **13(4)** (2004), pp. 907-929.

[3] Chopin, N. A Sequential Particle Filter Method for Static Models. *Biometrika* **89** (2002), pp. 539–551.

[4] Douc, R., Guillin, A., Marin, J.M., and Robert, C.P.: Convergence Of Adaptive Mixtures Of Importance Sampling Schemes. *The Annals of Statistics* **35** (2007), pp. 420-448.

[5] Fahs, F., Mittelhammer, R., and Yoder, J.: Generalized Method of Truncated Moments Estimation of Censored Equation Systems, Working Paper Series WP 2013-4, School of Economic Sciences, Washington State University, 2013.

[6] Geweke, J.: *Contemporary Bayesian Econometrics and Statistics.* Wiley Series in Probability and Statistics. John Wiley and Sons, 2005.

[7] Geweke, J., Keane, M. P., and Runkle, D.: Alternative Computational Approaches to Inference in the Multinomial Probit Model. *The Review of Economics and Statistics* **76** (1994), pp. 609–32.

[8] Huang, C., Sloan, F., and Adamache, K. (1987). Estimation of Seemingly Unrelated Tobit Regressions via the EM Algorithm. *Journal of Business and Economic Statistics* **5** (1987), pp. 425–430.

[9] Huang, R: Estimation of the SUR Tobit Model via the MCECM Algorithm, *Economics Letters* **64** (1999), pp. 25–30.

[10] Huang, R.: Bayesian Analysis of the SUR Tobit Model, *Applied Economics Letters* **8** (2001), pp. 617–622.

[11] Lynch, S.M.: *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* Springer, 2007.

[12] Qian, H. Estimating SUR Tobit Model while errors are gaussian scale mixtures: with an application to high frequency financial data, MPRA Paper No. 31509, available at http://mpra.ub.uni-muenchen.de/31509/ , 2009 .

[13] Tobin, J.: Estimation of relationships for limited dependent variables. *Econometrica* **26** (1958), pp. 24-36.

Numerical Methods for Option Pricing

Marco Cassader[1], Tomas Tichy[2]

Abstract. European Puts and Calls are known as vanilla options, that is, they are the most basic type of options, with relatively simple features and payoffs. Whilst for these options a closed form solution exists, in the global financial markets are traded different contingent claims with more complicated payoffs. Many models are used to price European options and it is difficult to define the best one. In fact, for different strike prices and in different time periods one model is preferred to others since it is more suitable to find an option value close to the market price. Based on the concept of risk neutral pricing and using stochastic calculus, variety of numerical models and methodologies have been used to determine the theoretical value of an option. In this paper, we discuss four different techniques used for pricing options and we test them with some empirical analysis.

Keywords: Numerical techniques, options pricing, finite difference methods.

JEL Classification: C44
AMS Classification: 35Q99

1 Introduction

The aim of this work is to present and evaluate the benefit of a numerical procedure for option price: the finite difference methods. In particular, we analyze some significant improvement in the grid generation ([3] or [5]) and we discuss the results of an empirical application. Since the introduction of Fischer Black and Myron Scholes' famous option pricing model [2] in 1973, several authors have proposed alternative models and methodologies for pricing options accurately ([6],[9],[1] or [8]). In this work, we propose a comparison between three finite difference methods and the Black-Scholes formula to price some European option written on a stock index. In particular, finite difference methods are generally characterized by the construction of a time and space grid structure as described in [5] and [10]. Then numerical methods are used to estimate the parabolic partial differential equation of the Ito formula for option pricing. One central point in the finite different method is the grid construction created with the aim to guarantee the existence of a solution in each time and space step. In this work we deal with this problem proposing a solution that take into account the maturity of the option and the possible price steps.

In Section 2 we present the different methods used for the options price and in Section 3 we discuss the results of an empirical application.

2 The finite difference methods

In this section, we present three of the main finite difference techniques used for options pricing. In particular, finite difference (FD) methods are used to numerically approximate the solutions of certain ordinary and partial differential equations (PDE). In the case of a bivariate parabolic PDE, such as

$$dC(S_t,t) = \frac{\partial C}{\partial S}dS_t + \frac{\partial C}{\partial t}dt + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}(dS_t)^2 = \frac{\partial C}{\partial S}(\mu\, S_t\, dt + \sigma\, S_t\, dB_t) + \frac{\partial C}{\partial t}dt + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}(\sigma^2\, S_t^2\, dt), \quad (1)$$

we start by establishing a rectangular solution domain in the two variables, S and t. We then form finite difference approximations to each of the derivative terms in the PDE (see Tavella [11]). FD methods create a mathematical relationship which links together every point on the solution domain, like a chain. The first links in the chain are the boundary conditions and from these, we 'discover' what every other point in the domain has to be. The most popular FD methods used in computational finance are: implicit Euler, explicit Euler and the Crank-Nicolson method. Using each of these three methods has its advantages and disadvantages. The easiest scheme of the three popular FD methods to implement is the explicit Euler method. Implicit Euler and Crank-Nicolson are implicit methods, which generally require a system of linear equations to be solved at each time step, which can be computationally intensive on a fine mesh. The main disadvantage to using explicit Euler is that it is unstable for certain choices of domain discretization. Though implicit Euler and Crank-Nicolson involve solving

[1] University of Bergamo, Via dei Caniana 2, Bergamo, Italy, email: marco.cassader@unibg.it and VSB-TUO, Sokolská 33, 701 21 Ostrava, Czech Republic
[2] VSB-TUO, Sokolská 33, 701 21 Ostrava, Czech Republic, e-mail: tomas.tichy@vsb.cz

linear systems of equations, they are each unconditionally stable with respect to the domain discretization. Crank-Nicolson exhibits the greatest accuracy of the three for a given domain discretization. The reason that Finite Difference methods (or FD) are a popular choice for pricing options is that all options will satisfy the Black-Scholes PDE, (1) or appropriate variants of it. The difference between each option contract is in determining the boundary conditions that it satisfies. Finite Difference (or FD) methods can be applied to American (early exercise) options and they can also be used for many exotic contracts.

The general differential equation that options must satisfy is the following:

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf \tag{2}$$

The aim is to construct a grid discretizing time and space. In particular, suppose that the life of the option is T. We divide this into N equally spaced intervals of length $\Delta t = T/N$. A total of $N+1$ times are therefore considered: $0, \Delta t, 2\Delta t, \dots, T$. Then, suppose that S_{max} is a stock price sufficiently high that, when it is reached, the put has virtually no value. We define $\Delta S = S_{max}/M$ and consider a total of $M+1$ equally spaced stock prices: $0, \Delta S, 2\Delta S, \dots, S_{max}$. The level S_m, is chosen so that one of these is the current stock price. The time points and stock price points define a grid consisting of a total of $(M+1)(N+1)$ points. The (i, j) point on the grid is the point that corresponds to time $i\ \Delta t$ and stock price ΔS. We will use the variable $f_{i,j}$ to denote the value of the option at the (i, j) point.

In the next subsections we describe the mathematical formulation of the difference methods and we derive the parameters to construct the optimal grid.

2.1 Implicit finite difference method

The implicit method allows to use a large number of *S*-mesh points without taking ridiculously small time-steps. We can solve our system of linear equations using the LU decomposition method and the use of this technique makes implicit method as almost as efficient as the explicit method in terms of arithmetical operations per time-step. As fewer time-steps need to be taken, the implicit finite difference method, which is unconditionally stable, is more efficient over-all than the explicit method.

In particular, for an interior point (i, j) on the grid we have to approximate $\frac{\partial f}{\partial t}, \frac{\partial f}{\partial S}, \frac{\partial^2 f}{\partial S^2}$. For the second term we can use a forward difference approximation: $\frac{\partial f}{\partial S} = \frac{f_{i,j+1} - f_{i,j}}{\Delta S}$ or we can consider a backward difference: $\frac{\partial f}{\partial S} = \frac{f_{i,j} - f_{i,j-1}}{\Delta S}$. The better solution consist of a symmetrical approximation averaging the previous two:

$$\frac{\partial f}{\partial S} = \frac{f_{i,j+1} - f_{i,j-1}}{2\Delta S} \tag{3}$$

Then, for the first term we will use a forward difference approximation:

$$\frac{\partial f}{\partial t} = \frac{f_{i+1,j} - f_{i,j}}{\Delta t} \tag{4}$$

Consider next $\frac{\partial^2 f}{\partial S^2}$. The forward and backward difference approximations for $\frac{\partial f}{\partial S}$ at the (i, j) point are given by the previous equations. Hence a finite difference approximation for $\frac{\partial^2 f}{\partial S^2}$ at the point (i, j) is:

$$\frac{\partial^2 f}{\partial S^2} = \frac{\left(\frac{f_{i,j+1} - f_{i,j}}{\Delta S} - \frac{f_{i,j} - f_{i,j-1}}{\Delta S}\right)}{\Delta S} \tag{5}$$

$$\frac{\partial^2 f}{\partial S^2} = \frac{f_{i,j+1} + f_{i,j-1} - 2f_{i,j}}{\Delta S^2} \tag{6}$$

Substituting the equations in the first one and nothing that $S = j\Delta S$ gives:

$$\frac{f_{i+1,j} - f_{i,j}}{\Delta t} + rj\Delta S \frac{f_{i,j+1} - f_{i,j-1}}{2\Delta S} + \frac{1}{2}\sigma^2 j^2 \Delta S^2 \frac{f_{i,j+1} + f_{i,j-1} - 2f_{i,j}}{\Delta S^2} = rf_{i,j} \tag{7}$$

for $j = 1,2, \dots, M-1$ and $i = 0,1, \dots, N-1$. Rearranging terms, we obtain:

$$a_j f_{i,j-1} + b_j f_{i,j} + c_j f_{i,j+1} = f_{i+1,j} \tag{8}$$

where

$$\begin{aligned} a_j &= \tfrac{1}{2} rj\Delta t - \tfrac{1}{2}\sigma^2 j^2 \Delta t \\ b_j &= 1 + \sigma^2 j^2 \Delta t + r\Delta t \\ c_j &= -\tfrac{1}{2} rj\Delta t - \tfrac{1}{2}\sigma^2 j^2 \Delta t \end{aligned} \tag{9}$$

The value of the put at time T is $max(K - S_T, 0)$, where S_T is the stock price at time T. Then, we evaluate the price at the boundary points:

$$f_{N,j} = \max(K - j\Delta S, 0) \quad j = 0,1, \dots, M \tag{10}$$

$$f_{i,0} = K \quad i = 0,1, \dots, N \tag{11}$$

and assuming that the put option is worth zero when $S = S_{max}$:

$$\mathrm{f_{i,M}} = 0 \quad \mathrm{i} = 0,1, \dots, \mathrm{N} \tag{12}$$

It remains to use the computed equation to arrive at the value of f at all other points. First the points corresponding to time $T - \Delta t$ are tackled. With $i = N - 1$ it gives:

$$\mathrm{a_j f_{N-1,j-1} + b_j f_{N-1,j} + c_j f_{N-1,j+1} = f_{N,j}} \tag{13}$$

For $j = 1,2, \dots, M - 1$. The right-hand side of these equations are known from the previous one. Moreover we know that

$$\begin{aligned} \mathrm{f_{N-1,0}} &= \mathrm{K} \\ \mathrm{f_{N-1,M}} &= 0 \end{aligned} \tag{14}$$

These are $M - 1$ simultaneous equations that can be solved for $M - 1$ unknown parameters. After this has been done, each value of $f_{N-1,j}$ is compared with $K - j\Delta S$ and we evaluate the possibility to an early exercise. The implicit finite difference method has the advantage of being very robust. It always converges to the solution of the differential equation as ΔS and Δt approach zero.

2.2 Explicit finite difference method

Described by Brennan and Schwartz [4] and developed by Hull and White [7] and Boyle and Tian [3], the explicit finite difference method is very popular. The main reason is its simplicity to program. Considering the equation (2), if the values of $\frac{\partial f}{\partial S}$ and $\frac{\partial^2 f}{\partial S^2}$ at point (i, j) on the grid are assumed to be the same as at point $(i + 1, j)$, we can rewrite it with the following definitions:

$$\begin{aligned} \frac{\partial \mathrm{f}}{\partial \mathrm{S}} &= \frac{\mathrm{f_{i+1,j+1} - f_{i+1,j-1}}}{2\Delta \mathrm{S}} \\ \frac{\partial^2 \mathrm{f}}{\partial \mathrm{S}^2} &= \frac{\mathrm{f_{i+1,j+1} + f_{i+1,j-1} - 2f_{i+1,j}}}{\Delta \mathrm{S}^2} \end{aligned} \tag{15}$$

Thus, the difference equation is:

$$\frac{\mathrm{f_{i+1,j} - f_{i,j}}}{\Delta \mathrm{t}} + \mathrm{rj}\Delta \mathrm{S}\frac{\mathrm{f_{i+1,j+1} - f_{i+1,j-1}}}{2\Delta \mathrm{S}} + \frac{1}{2}\sigma^2 \mathrm{j}^2 \Delta \mathrm{S}^2 \frac{\mathrm{f_{i+1,j+1} + f_{i+1,j-1} - 2f_{i+1,j}}}{\Delta \mathrm{S}^2} = \mathrm{rf_{i,j}} \tag{16}$$

We can rewrite the previous formula in the following methodology:

$$\mathrm{f_{i,j} = a_j^* f_{i+1,j-1} + b_j^* f_{i+1,j} + c_j^* f_{i+1,j+1}} \tag{17}$$

where:

$$\begin{aligned} \mathrm{a_j^*} &= \frac{1}{1+\mathrm{r}\Delta \mathrm{t}}\left(-\frac{1}{2}\mathrm{rj}\Delta \mathrm{t} + \frac{1}{2}\sigma^2 \mathrm{j}^2 \Delta \mathrm{t}\right) \\ \mathrm{b_j^*} &= \frac{1}{1+\mathrm{r}\Delta \mathrm{t}}\left(1 - \sigma^2 \mathrm{j}^2 \Delta \mathrm{t}\right) \\ \mathrm{c_j^*} &= \frac{1}{1+\mathrm{r}\Delta \mathrm{t}}\left(\frac{1}{2}\mathrm{rj}\Delta \mathrm{t} + \frac{1}{2}\sigma^2 \mathrm{j}^2 \Delta \mathrm{t}\right) \end{aligned} \tag{18}$$

One problem with the explicit formula is that the constraints imposed upon the sizes of the ΔS steps can make it very inefficient. In order that none of the probabilities are negative, ΔS must be larger than $\lambda\sigma S\sqrt{\Delta t}$, where the constant coefficient λ, as suggested by Boyle and Tian [3] can be taken as $\sqrt{1.5}$. The S in this case must be taken as the largest possible S on the grid; it is for this reason that the dimensionless log-transformed version may be more efficient for the explicit method. For the log-transformed grid, where $x = ln(S)$, then δx must be larger than $\sigma\sqrt{\Delta t}$.

2.3 Crank-Nicholson finite difference method

Finally, Crank-Nicolson is computed as the average of the explicit and implicit methods. The advantage is the faster convergence than either the explicit or implicit method. The explicit scheme is given by (17) and the implicit by (13). We take the average of the two equations to get:

$$f_{i,j} + f_{i-1,j} = a_j f_{i-1,j-1} + b_j f_{i-1,j} + c_j f_{i-1,j+1} + a_j^* f_{i,j-1} + b_j^* f_{i,j} + c_j^* f_{i,j+1} \tag{19}$$

Let:

$$\eta_{i,j} = a_j^* f_{i,j} - f_{i,j-1} - b_j^* f_{i,j} - c_j^* f_{i,j+1} \tag{20}$$

We obtain the following linear problem to solve:

$$\eta_{i,j} = a_j f_{i-1,j-1} + b_j f_{i-1,j} + c_j f_{i-1,j+1} - f_{i-1,j} \tag{21}$$

Notice that in the Crank-Nicolson method one new point is defined considering six different points.

3 Empirical analysis with market data

In this section, we discuss the results of an empirical analysis considering stock index DAX options on 15 September 2011. In particular, we take 194 call and put options with the same maturity of 183 days. We focus our analysis on the comparison between the market price, market gamma and delta with the results of different methods: the classical Black-Scholes formula, the implicit method, the explicit method and the Crank-Nicholson.

We present the results using the realized volatility taking into account 1 year of historical observation of the underlying DAX stock index. Note that we compute our analysis considering a fixed time step (1 day) for all the methods and different space step for the explicit method. This is necessary as a consequence of the discussion done in the section 2.2. In particular, we let $S_{min} = 0$ and $S_{max} = 2 \times max(K(i), S)$ for $i = 1, \dots, 194$ for all the methods with the exception of the explicit one. In fact, in that case we take the values suggested in [3] Moreover, we use a uniform grid with the introduction of the approximation space step as described by the literature (see Boyle and Tian [3] and Tavella and Randall [7]).

In the following part of this section, we discuss the results of our pricing models considering not only the option price but also the computation of the Delta and Gamma Greeks. In fact, these components are very considered for risk management operators and they are a key point to hedge the options market risk.

Figure 1 shows the differences between the market price and the price value computed with the different methods for a given strike price. In particular the left side shows the result for the calls option pricing and the right part for the put prices. The lines represent a correct pricing process when their values are equal to zero, an overestimation in the case of negative value or an underestimation when for a given strike the curve shows positive value.

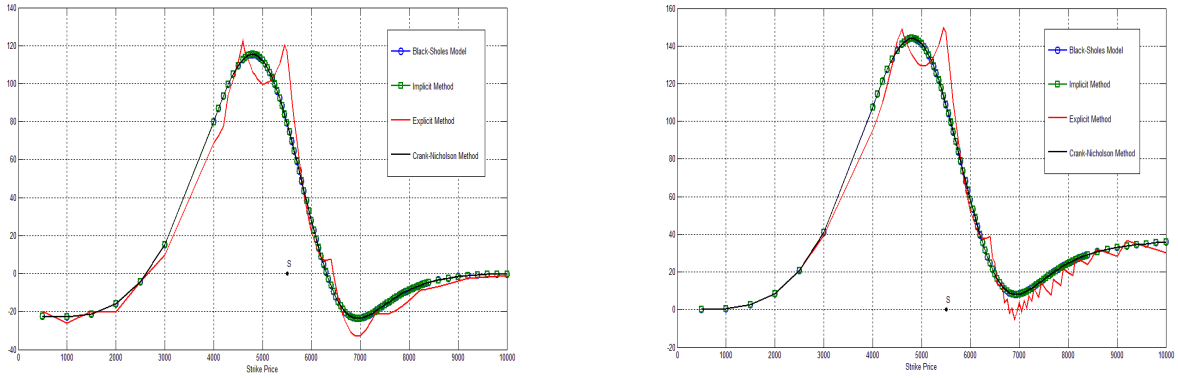


Figure 1 Differences between market option price and different methods for Call and Put

From the analysis of the Black-Scholes model and finite difference methods we deduce:

- In general, every model underestimate the market value for put options with different strike prices while they overestimate the market price of call options for deep in/out of the money strike prices.
- Black-Scholes formula, implicit and Crank-Nicholson methods show similar results for every strike price.
- Strike prices close to the underlying stock index show an inefficient estimation for all the methods while for deep out or deep in the money strike prices the pricing of the methods converge. This convergence is related

to the market price when we are pricing deep in the money options. However, for the out of the money cases we observe an underestimation of the call prices and an overestimation for the put prices.

- The options price methods work correctly only for few strike prices. In the other cases, we notice that the explicit finite difference method works better than the others for strike prices lower than the underlying value both in the calls and puts pricing process.
- Whilst, implicit, Crank-Nicholson and Black-Scholes methods show a regular behavior, the explicit method performs better than the others for some strike price and it presents the worst pricing for other values.
- For strike prices close to the underlying index value all the methods show a significant estimation differences with respect to the market price. This inefficiency could be justified by the need of a volatility process to specify the smile phenomena.

An important aspect to consider when we deal with the option pricing is the computation of the Greek letters. In particular, the analysis of the delta and gamma is a central problem for every risk management operator. Figures 2 and 3 illustrate the values of the delta and gamma for different strike prices and compare the market Greeks values with those derived from the Black-Scholes model and the finite difference methods. In both figures we report the separation between call and put options and we analyze the evolution for different strike prices. In particular, it is evident that in Figure 2:

- The value of Black-Scholes formula, implicit and Crank-Nicholson method is very similar. They work very well to estimate the Delta for at the money options and for in the money and deep in the money call options.
- Generally, every methods overestimate the Delta for call options while for out of the money put options only the explicit method show this behavior.
- Explicit method is clearly inefficient with respect to the other methodologies to estimate the Delta of call and put options.
- There is a problem for every method to estimate the Delta for strike price rather lower than the underlying index value. This phenomenon happens without a distinction between call and put options.

To recap, the computation of the delta using Black-Scholes formula and two finite difference methods (implicit and Crank-Nicholson) is optimal for strike prices close to the underlying value. In the other cases we observe some important differences in the estimation process with the exception of deep in the money call options. Moreover, it is evident that the explicit method does not give the same results as in the price estimation for every strike prices.

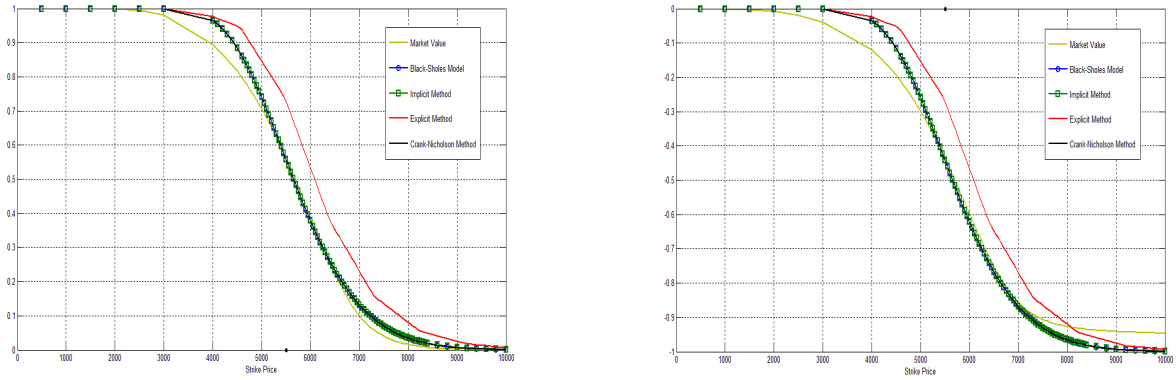


Figure 2 Delta Call and Put

The last analysis focuses on the computation of the gamma measure. Figure 3 shows the results of our study comparing the market gamma with the estimated values using different methods. Analyzing the left and right side of the Figure 3 we can deduce:

- There are not significant differences between Black-Scholes formula and all the finite difference methods.
- They, generally, underestimate the market gamma for deep in the money and deep out of the money call and put options. In the other case they overestimate the gamma values. In particular, we notice an efficient estimation for in the money calls and out of the money puts.
- The explicit method works better than the others for some strike price ranges. In particular for out of money call and out of the money put options.

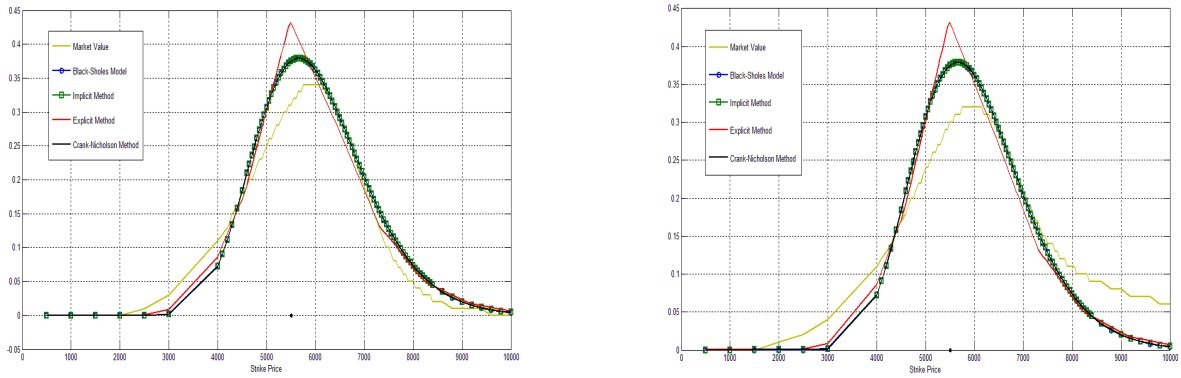


Figure 3 Gamma Call and Put

4 Conclusion

In this work we present four different methods to compute the price, the delta and the gamma of different call and put options written on the DAX stock index with the same maturity. In particular, we use the classical Black-Scholes model and we introduce three finite different methods with a specific grid construction: the implicit, the explicit and the Crank-Nicholson methods. The results of our analysis suggest to introduce different methodologies to obtain the better estimation of the market value with respect to the strike price taken into account. However, a better estimation and a time evolution represent future developments of our work with potential application to exotic options.

Acknowledgements

The research was supported through the Czech Science Foundation (GAČR) under project 13-13142S and furthermore by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and the European Social Fund (CZ.1.07/2.3.00/20.0296) and SP2014/16, an SGS research project of VSB-TU Ostrava. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005) is also appreciated.

References

[1] Andersen, L., Andreasen, J.: Jump-diffusion processes: Volatility smile fitting and numerical methods for option pricing. *Review of Derivatives Research* **4** (2000), 231-262.
[2] Black, F., and Scholes, M.: The pricing of options and corporate liabilities. *Journal of Political Economy* **81** (1973), 637-654.
[3] Boyle, P. P., and Tian, Y.: An explicit finite difference approach to the pricing of barrier options. *Applied Mathematical Finance* **5** (1998), 17-43.
[4] Brennan, M. J., Schwartz, E. S.: Finite difference methods and jump processes arising in the pricing of contingent claims: A synthesis. *Journal of Financial and Quantitative Analysis* **13** (1978), 461-474.
[5] Duffy, D. J.: Finite Difference methods in financial engineering: a Partial Differential Equation approach. John Wiley & Sons, New York, 2006.
[6] Ekström, E., Lötstedt, P., Sydow, L. V., Tysk, J.: Numerical option pricing in the presence of bubbles. *Quantitative Finance* **11** (2011), 1125-1128.
[7] Hull, J., White, A.: Valuing derivative securities using the explicit finite difference method. *Journal of Financial and Quantitative Analysis* **25** (1990), 87-100.
[8] Pooley, D. M., Forsyth, P. A., Vetzal, K. R.: Numerical convergence properties of option pricing PDEs with uncertain volatility. *IMA Journal of Numerical Analysis* **23** (2003), 241-267.
[9] Rachev, S., Mittnik, S.: Stable Paretian models in finance, Wiley, New York, 2000.
[10] Tavella, D., Randall, C.: *Pricing financial instruments: The finite difference method*. Chichester: John Wiley & Sons, New York, 2000.
[11] Tavella, D.: Quantitative methods in derivatives pricing: an introduction to computational finance. John Wiley & Sons, New York, 2003.

Generating Data Model for the Public Service System Design from the OpenStreetMap: Network and the Shortest Paths

Matej Cebecauer[1]

Abstract. When designing public service systems, such as emergency health-care, police or fire brigades, we need to know demand of customers, their locations, possible locations of service centers and the traveling distances between them. Traveling distances are mostly approximated by the length of shortest paths, which need to be computed in advance. In recent years, the availability of open data is rapidly growing. Remarkable example is the OpenStreetMap (OSM) portal providing detailed geographical information. We use OSM data to extract the road network and to identify customers' locations. For example, for the area of the Slovak Republic we obtained a network representing customers interconnected via the road network composed from nearly 2 million nodes. In this contribution, we propose an efficient algorithm to compute the shortest paths which is based on the hierarchical decomposition where smaller graph segments are interconnected through the macroscopic network. We evaluate the proposed algorithm by comparing it with the traditional shortest path algorithm.

Keywords: shortest path, decomposition, public service system.

JEL classification: C44
AMS classification: 90C15

1 Introduction

The public service systems such as emergency health-care, police or fire brigades serve customers mainly through road network. When we are designing these systems we need to know demand of customers, their locations, possible locations of service centres and the traveling distances between them. These four parts represent the data model of the location problem.

The geographical area of designed public service system could be spatially large with a lot of customers. Most of the problems related to the facility location are known to be NP-hard [2], therefore they cannot be computed in reasonable time especially for large instances of optimization problems. It may be impossible to consider every individual customer in the model. Aggregation approaches are valuable tools to obtain manageable models [8]. The result of aggregation is a smaller number of aggregated customers representing original unaggregated customers. There is a strong stream of literature in location science studying aggregation methods and corresponding errors, for comprehensive overview see [4] and references therein.

We are interested in the location problems with hundreds of thousands of customers where it is not possible to used complete distance matrix because of a memory limitations. There is possibility to use complete road network and find the shortest path between two nodes when it is requested. The shortest path could be found by traditional algorithms [1, 3]. The time necessary to find the shortest path arises with the size of the road network as well as the memory usage. To deal with this consequences, graph decomposition proposed in [5, 6, 7] is helpful.

In this paper we propose a decomposition method inspired by [5], but we are not searching for an ideal decomposition. We decompose the road network using municipality boundaries to the segments with

[1]University of Žilina, Department of Transportation Networks, Univerzitna 8215/1 Zilina 01026 Slovakia, matej.cebecauer@gmail.com

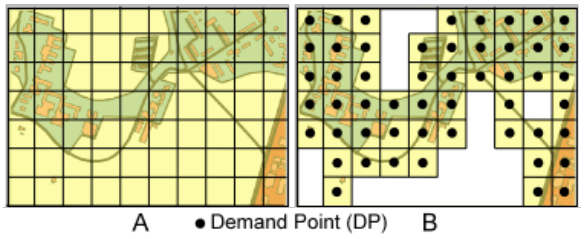


Figure 1 Illustration of generating DPs

smaller road networks. The experiments are computed on the very big network with almost 2 million nodes.

The paper is organized as follows: section 2 introduces the data model. In section 3, we describe decomposition method of original road network and in section 4 we propose the shortest path algorithm. Results of computational experiments are reported in section 5. To conclude, we summarize our findings in section 6.

2 Process of Generating Data Model of the Road Network

In this section we describe how the data model of the road network is generated. Resulting data model consists from the road network and Demand Points (DP) that represents customers. The OSM contains all necessary data necessary to generate DP and to extract road network. We are especially interested in layers of buildings, roads, residential, industrial and commercial areas. We are using these layers to model a potential presence of population in spatial area and to decide where DPs can be situated.

We generate DPs with a simple rule based on the OSM layers that we discussed above. First of all we generate spatial grid, which consists of the uniform square cells with a size 100 meters. Each cell contains parts from OSM layers which are situated inside these cells. After application of the spatial grid, DPs are located as centroids of cells. However not all cells are ideal to be DPs. Only the cells that collide with same data from OSM layers are consider to have representative DPs. The process of generating DPs is visualized in Figure 1.

Finally, the all DPs are connected to the road network by creating new node in a way that minimizes the distance of connection to the road network. The Slovak Republic has 1 966 092 nodes (including 666 824 DPs) and 2 090 807 edges. The distance matrix of 666 824 DPs can easily exceed pc memory. Alternative to the distance matrix could be the complete road network but consequence is that it takes more time to achieve the shortest path from graph. In further subsections we propose alternative approach to the complete road network.

3 Graph Decomposition Method

In this section we propose decomposition method inspired by [5], but we are not looking for a ideal decomposition of graph. We use OSM layer containing boundaries of administrative zones (AZ) to decompose the complete road network. In the OSM, there are included different levels of administrative decompositions, see Figure 2. Especially the AZs of cities and towns perfectly approximate clusters, with bigger density roads inside and low number of transitions between adjacent zones.

If the graph is decomposed, the possibility to find the shortest path needs to be preserved. We propose a technique, which finds the shortest path through the decomposed graphs. We are using a macroscopic graph, which consists only from a gateway of all decomposed graphs. These gateways are nodes that symbolize a transition between two adjacent AZs. The decomposition is visualized in Figure 3A-D. A transition node is selected from a pair of road nodes where one is in or out of AZ and the next node in the order has different location statement as a previous node (Figure 3C). The first node in the order of this pair is considered as the transition node. It is important to note that in the OSM, each road is defined by the order of nodes, which guarantee that if two adjacent AZs have some transition road then the transition node is the same in both of them (Figure 3C-D).

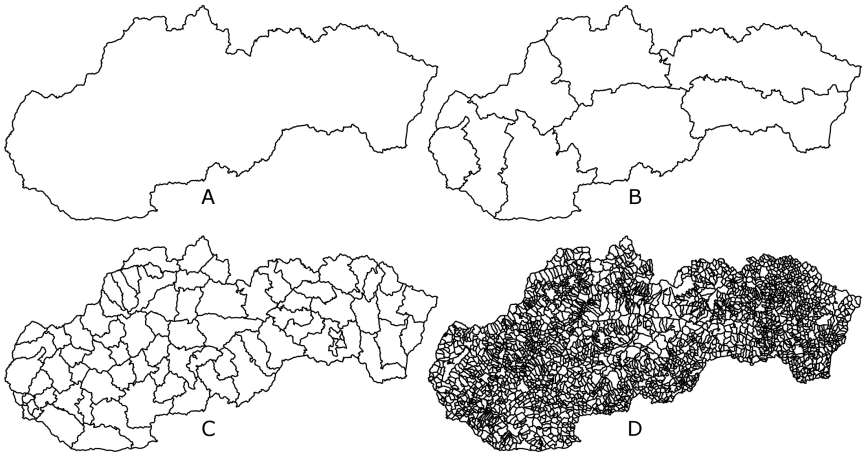


Figure 2 Administrative boundaries: (A) Slovak Republic boundary, (B) Regions administrative boundaries, (C) Districts administrative boundaries, (D) Cities and towns administrative boundaries.

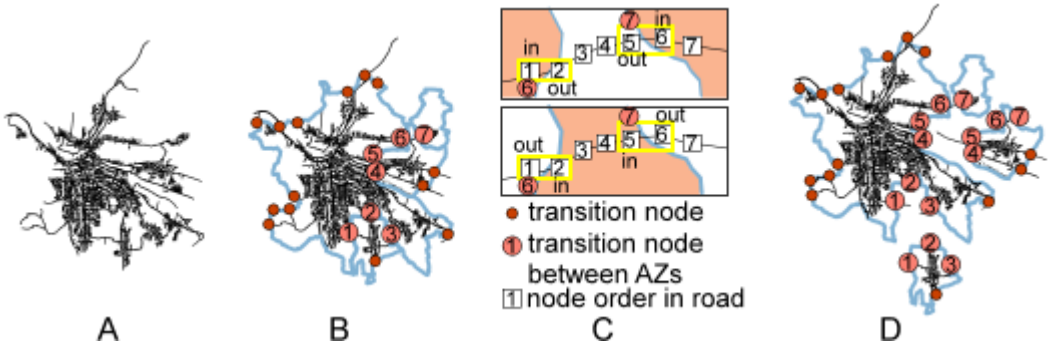


Figure 3 Illustration of graph decomposition on the small geographical area of city Žilina: (A) road network of Žilina and its surroundings, (B) intersections, of towns and villages administrative boundaries with road network, is visualized with transition nodes, (C) description of a transition node selection technique from the road network, (D) illustration of decomposition where the each transition nodes between AZs is included in both adjacent AZs.

To generate macroscopic graph, we only need the transition nodes. The original graph G with 1 966 092 nodes is decomposed to m subgraphs defined by AZs (see Figure 4A). For each subgraphs $k = 1 \dots m$ we compute the distances matrix of transition nodes D_k from which we generate the distance graph G_k (see Figure 4B). Further we combine all G_k for $k = 1 \dots m$ (illustrated in Figure 4C) to the one macroscopic graph (see Figure 4D). For example, the macroscopic graph of entire Slovak Republic is significantly smaller and has only 8072 nodes which is 0,04% of the original graph.

In next section we propose the shortest path algorithm which is using the macroscopic graph to find the shortest path of two random nodes of the original graph.

4 The Shortest Path Algorithm

In previous section, we introduced decomposition and macroscopic graph. The macroscopic graph does not allow to find the shortest path between regular nodes from original graphs but only between transition nodes. In this section, we propose the algorithm which uses the macroscopic graph to find the shortest path between arbitrary nodes of original graph.

The algorithm generates a new smaller graph G_a that uses the knowledge of transition nodes that represent gateways between AZs. The AZ of node k is denoted as $AZ(k)$ and transition nodes of $AZ(k)$ as the $N_{AZ(k)}$. There exist two basic scenarios, if we are finding path between a start node s and end node e: the $AZ(s)$ and $AZ(e)$ could be different AZs or the same AZ. In both cases G_a is constructed as follows:

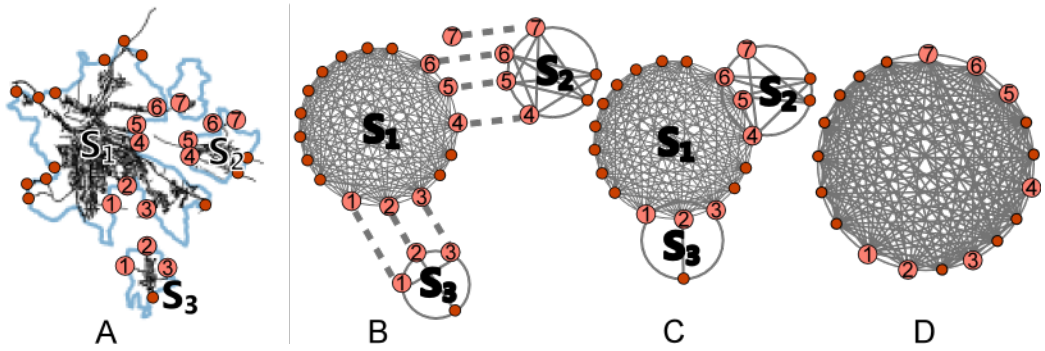


Figure 4 Visualization of generating macroscopic graph: (A) AZs with complet road network (B) Subgraphs derived from distance matrices between transition nodes, (C) Illustration of subgraphs combination, (D) Final macroscopic graph consisting from all transition nodes.

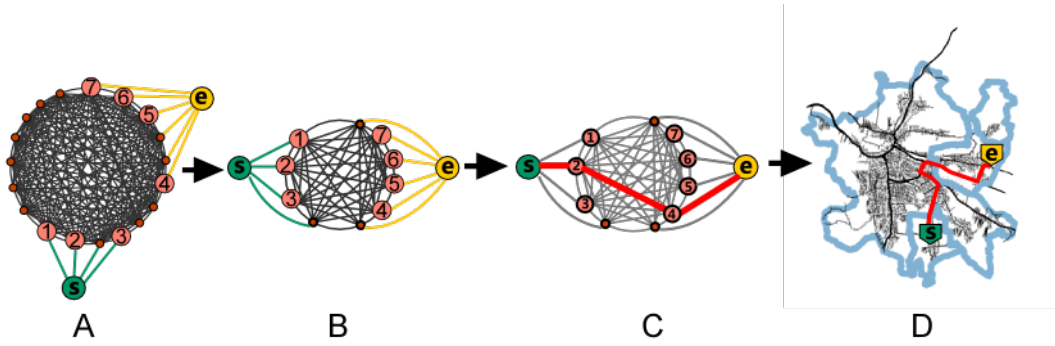


Figure 5 Illustration of the shortest path algorithm, which is using macroscopic graph in regard to the scenario, where s and e are in the different AZs: (A) Visualization of STEP 1 and 2, where the start node s and the end node e are connected to the transition nodes with distance obtained from the shortest path on the subgraphs $AZ(s)$ and $AZ(e)$ of original graph. (B) Illustration of STEP 3 where subgraph is separated from macroscopic graph defined only by all transition nodes $N_{AZ(s)}$ and $N_{AZ(e)}$. (C,D) Visualization of the shortest path between nodes s and e.

STEP 0. **:** Initialize empty graph G_a.
STEP 1. **:** Add the s and e nodes to the G_a.
STEP 2. **:** Add transition nodes of $N_{AZ(s)}$ and $N_{AZ(e)}$ to G_a . Further add edges consisting from the shortest path obtained from subgraph of $AZ(s)$ between $N_{AZ(s)}$ and node s to G_a. Apply similar process for node e and $N_{AZ(e)}$ if $AZ(s) \neq AZ(e)$.
STEP 3. **:** Add edges from the macroscopic graph among all transition nodes of $N_{AZ(s)}$ and $N_{AZ(e)}$ to G_a .

The steps above are the same in both cases, but if the s and e nodes are in the same AZ then also the shortest path between s and e in the AZ needs to be included to ensure the optimality of the shortest path. Further we have added additional step to the algorithm:

STEP 4. **:** If $AZ(s) = AZ(e)$ than add edges with distance of the shortest path, obtained from the subgraph of $AZ(s)$ between s and e.

STEP 5. **:** Finally, the shortest path between nodes s and e in the graph G_a is founded by one of the traditional algorithms.

The algorithm is visualized in the Figure 5A-D for the case if s and e are in the different AZs and if they are in the same AZ is illustrated in the Figure 6A-D.

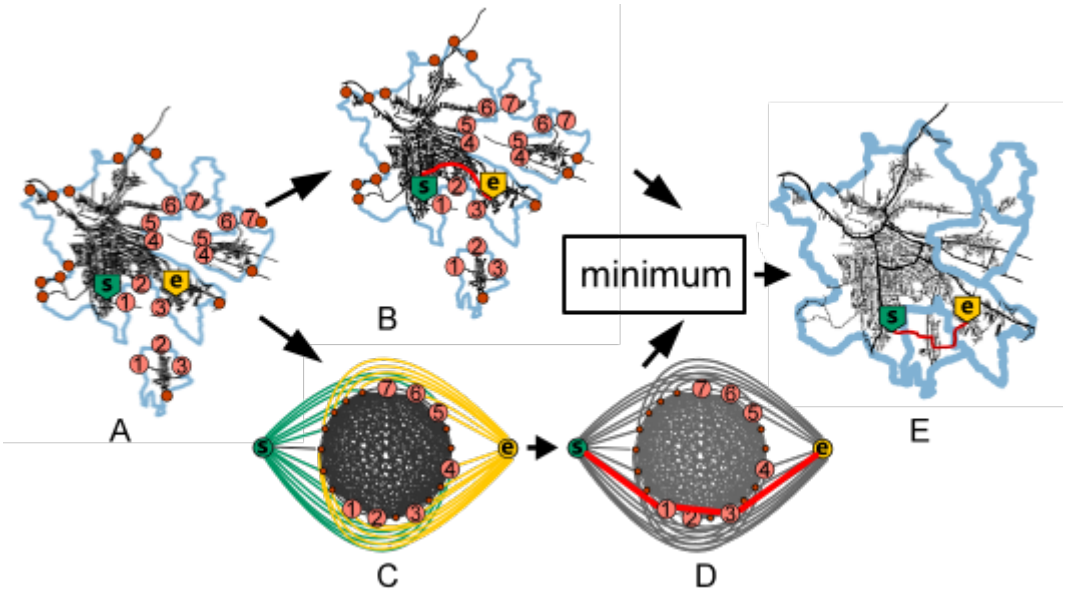


Figure 6 Illustration of the shortest path algorithm, which is using macroscopic graph in regard to scenario where s and e are in the same AZ: (A) Visualization of nodes s and e in the AZ (B) Illustration of STEP 4 with highlighted the shortest path between nodes s and e in the AZ. (C) Visualization of STEP 1,2 and 3 where the start node s and the end node e are connected to the transition nodes with distance obtained from the shortest path on the subgraphs $AZ(s)$ and $AZ(e)$ of the original graph. (D) Highlighted shortest path in the graph G_a (E) Visualization of the shortest path between nodes s and e chosen as a minimum from B and D.

areas	number of nodes	number of edges	number of AZs	number of transition nodes
Žilina region	261 057	274 571	315	918
Slovak Republic	1 966 092	2 090 807	2929	8 072

Table 1 Statistic of geographical areas

5 Computational Experiments

In this section, we evaluate the time and memory efficiency of the method proposed above. Further we compare this algorithm with algorithm [1], which is utilized for the shortest path only between two nodes. We compute experiments on a one smaller network of region Žilina and the entire Slovak Republic (see Table 1).

We randomly selected 10 000 pairs of nodes, where we measured the elapsed times and the memory demands. The statistics of experiments are summarized in the Tables 2 ,3. The decomposition algorithm has time and memory effectiveness significantly better. It is important to note that initialization, which is using subgraphs and macroscopic graph, takes time from 0.02-10 seconds with average value 0.873 seconds with standard deviation 1.132 seconds. The times in the Table 3 could be achieved only if all data are in the pc memory.

areas	minimal	average	maximal time	standard deviation	memory usage
Žilina region	1.047 s	1.099 s	1.328 s	0.03 s	550 - 600 MB
Slovak Republic	8.610 s	8.854 s	10.313 s	0.155 s	3 - 3,1 GB

Table 2 Statistic of Dijkstra algorithm experiments

areas	minimal time	average time	maximal time	standard deviation	memory usage
region Žilina	<0.001 s	<0.001 s	0.005 s	<0.001 s	170 - 280 MB
Slovak Republic	<0.001 s	0.0012 s	0.0156 s	0.0018 s	170-300 MB

Table 3 Statistic of the decompositon algorithm experiments

6 Conclusions

From the results we can conclude that the method proposed in this paper is more independent from the graph size. The time that is needed to find the shortest path is almost the same in the small region Žilina as well as on the much more bigger graph of the entire Slovak Republic. This method could be used in two versions. If the memory demand is too big then proposed method can use files and load only important data to find the shortest path. If the pc memory demand is enough for all data then it is possible to achieve rapidly better time efficiency.

The decomposition method could find almost 9000 shortest path until traditional algorithm found only the one shortest path. We have proven that the decomposition algorithm could be a valid substitution of the distance matrix where only specific shortest path between two nodes is expected. In the situation, where is "one to many" or "many to many" requested the Dijkstra algorithm will be more efficient. Actually we do not think that utilized version of decomposed algorithm for "one to many" or "many to many", achieved better time efficiency in comparison to the Dijkstra algorithm. Utilized version of proposed method or not, it could be only way how to find shortest path if the complete graph with Dijkstra algorithm exceeds pc memory.

Acknowledgment

This work was supported by the research grants VEGA 1/0339/13 "Advanced microscopic modelling and complex data sources for designing spatially large public service systems" and APVV/-0760/-11 "Designing Fair Service Systems on Transportation Networks".

References

[1] Dijkstra, E, W.: A note on two problems in connexion with graphs *Numerische Mathematik* **1** (1959) 269-271

[2] Eiselt, H, A, and Marianov V.: *Foundations of Location Analysis*, Springer Science+Business Media, New York, (2011) 3–22.

[3] Floyd, R, W.: Algorithm 97: Shortest Path *Communications of the ACM 5* **6** (1962) 345

[4] Francis, R, I, and Lowe, T, J, and Rayco, M, B, and Tamir.: Aggregation error for location models: survey and analysis. *Annals of Operations Research* **167** (2009), 171–208

[5] Hu, T, C.: A decomposition Algorithm for Shortest Paths in a Network *Operations Research* **16** (1968), 91–102

[6] Land, A, H. and Stairs, S.: The Extension of the Cascade Algorithms to Larger Graphs *London School of Economics* **LSE-TNT-20** (1965)

[7] Murchland, J, D.: A New Method for Finding all Elementary Paths in a Complete Directed Graph *London School of Economics* **LSE-TNT-22** (1965)

[8] Rogers, D, F, and Plante, R, D.: Aggregation and disaggregation techniques and methodology in optimization *Operations Research* **39** (1991), 553–583

Optimal Scheduling of Vehicles for Wheelchair Users in Public Transport

Anna Černá[1], Jan Černý[2], Jaroslav Matuška[3]

Abstract. Public transportation operates on the network according to the given timetable. I.e. the set of regular vehicle journeys is given. Each journey is determined by vertices it passes and by departure and arrival times. However, only a minor part of vehicles are wheelchair friendly (abb. WFV). During one day, a WFV can be assigned to a sequence of journeys called a daily duty of the WFV.
The demand of a wheelchair users (WCU) is called covered by a wheelchair friendly vehicle (WFV) journey without transfer, if a WFV is assigned to the journey and the journey connects the origin and the destina-tion of the WCU in the demanded time. It is called covered by a pair of WFV jour-neys with one transfer if the first journey of the pair connects the origin of the WCU with the transfer stop and the second journey of the pair connects the transfer stop with the destination of the WCU, both in the demanded time.
The problem P1 is to find the minimum set of WFV daily duties covering the demand of all WCU's. If such a solution does not exist, then the problem P2 is to find such a set of daily duties of all WFV's that it covers the demand of maximal number of WCU's.
The paper proposes heuristic methods for the solution of problems P1 and P2.

Keywords: wheelchair user (WCU), WCU friendly vehicles, public transport, optimization, method.

JEL Classification: C61, C65, D63
AMS Classification: 05C35, 90B06

1 Introduction, the basic problem (BP)

Movement of disabled persons (e.g., visually impaired people or wheelchair users) is an important issue. In the scientific literature it is dealt from several aspects. [1] exerts effort to determine the extent to which Electric Indoor/Outdoor Powered wheelchair (EPIOC) users travel in vehicles, their concerns about safety, any accidents occurring during transportation and difficulties with their equipment. In [5] the development of digital map for these user groups is discussed and route planning of barrier-free routes with additional information is emphasized. In [2] it is shown how to elaborate design criteria for an integral accessible public area for all people including disabled persons.

This paper is focused on the following situation: A network, i.e. a unoriented graph $G = (V, E)$ is given. The set of vertices V represents the stops or stations where the set B of wheelchair users (briefly WCU) may board or get off the vehicles of public transport (trains, trams, buses etc.). In the sequel, the word 'stop' will be also used for station. The set of edges E represents segments of roads (in a general sense) between the stops. Each WCU $b \in B$ has the origin o_b, the destination s_b and further, either the departure time t_b [min] from o_b or the arrival time a_b [min] to s_b. Therefore, the basic form of WCU demand is the matrix

$$Q = \begin{pmatrix} o_1 & t_1 & s_1 & a_1 \\ o_2 & t_2 & s_2 & a_2 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

where each row $q_b = o_b, t_b, s_b, a_b$ corresponds to one WCU $b \in B$. The matrix Q has the property that in each row q_b, either the arrival time is undefined which will be written $a_b = 1440$ (i.e. the passenger wants to arrive before the midnight), or the departure time is undefined which will be written $t_b = 0$ (i.e. the passenger wants to depart after the midnight). All time data are in minutes, e.g. 6:17 = 6h + 17 min = 377 min.

[1] Vysoká škola ekonomická, FM, Jarošovská 1117, Jindřichův Hradec, Czech Republic, cerna@fm.vse.cz.
[2] Vysoká škola ekonomická, FM, Jarošovská 1117, Jindřichův Hradec, Czech Republic, cerny@fm.vse.cz.
[3] Univerzita Pardubice, DF JP, Studentská 95, Pardubice, Czech Republic, Jaroslav.Matuska@upce.cz.

Public transportation operates on the network G according to the given timetable. I.e. the set of regular vehicle journeys J is given. Each journey $j \in J$ has its departure stop v_{j0}, intermediate stops $v_{j1}, \ldots, v_{j,m(j)-1}$ and the final stop $v_{jm(j)}$ where $m(j) + 1$ is the number of stops of the journey j. Moreover, it has its departure time t_{j0} from v_{j0}, arrival and departure time a_{jk} and t_{jk} to and from the k^{th} stop of the journey j for each $k = 1, \ldots, m(j) - 1$ and the arrival time $a_{jm(j)}$ to the final stop. The fact that a vertex v is a stop (departure, intermediate or final) of the journey j will be denoted $v \in j$. A pair $i, j \in J$ is said w_{ij}-passable and denoted i ($w_{ij}\rightarrow$) j if there exist indices k, l such that $w_{ij} = v_{ik} = v_{jl}$ is the transfer stop and $a_{ik} < t_{jl}$. *A* sequence $\vec{j} = j(1), \ldots, j(r)$ of journeys is said ($w_{j(1)j(2)}$, …, $w_{j(r-1)j(r)}$)-passable if

$$j(1)\ (w_{j(1)j(2)}\rightarrow)\ j(2), \ldots, j(r-1)\ (w_{j(r-1)j(r)}\rightarrow)\ j(r)$$

This property enables a passenger to use the passable pair or sequence of journeys transferring from the previous journey to the next one at the common stop.

However, only a minor part of vehicles are wheelchair friendly (abb. WFV) i.e. enabling wheelchair loading and unloading. The set of all available WFV is denoted F_0 and their number is denoted $|F_0|$. Let us suppose that at the beginning (i.e. early morning), the WFV f is located at the stop g_f, $f \in F_0$.

A sequence $\tilde{j} = j(1),\ldots, j(r)$ is said a daily duty if the final stop of $j(n)$ equals to the departure stop of $j(n + 1)$ and the arrival time to the final stop of $j(n)$ is less than the departure time from the initial stop of $j(n + 1)$ for each $n = 1, \ldots, r$.

If $\tilde{j} = j(1),\ldots, j(r)$ is a daily duty and the sequence $j(0)\tilde{j} = j(0), j(1),\ldots, j(r)$ is also a daily duty, then the daily duty $j(0)\tilde{j}$ is said simple left extension of $\tilde{j}$. Similarly, if a sequence $\tilde{j}j(n+1) = j(1),\ldots, j(r), j(n+1)$ is also a daily duty, then the daily duty $\tilde{j}j(n+1)$ is said simple right extension of $\tilde{j}$.

A sequence $\tilde{j}_f = j_f(1),\ldots, j_f(r_f)$ is said a daily duty of the WFV f if the departure stop of the journey $j_f(1)$ is g_f. If a non empty set $F \subset F_0$ then the set of daily duties for all $f \in F$ is denoted $\tilde{J}_F$ and the set of all journeys belonging to the duties from $\tilde{J}_F$ is denoted J_F.

A sequence $\vec{j} = j(1), \ldots, j(r)$ is said WCU($w_{j(1)j(2)}$, …, $w_{j(r-1)j(r)}$)-passable by $\tilde{J}_F$ if it is $j(1) \in J_F, \ldots, j(r) \in J_F$ and $\vec{j}$ is ($w_{j(1)j(2)}$, …, $w_{j(r-1)j(r)}$)-passable.

A WCU $b \in B$ is said covered by a WCU($w_{j(1)j(2)}$, …, $w_{j(r-1)j(r)}$)-passable sequence $\vec{j} = j(1), \ldots, j(r) \in \tilde{J}_F$ if the following conditions hold: $o_b = v_{j(1)l}$, $t_b < t_{j(1)l}$, $l < k$ for $w_{j(1)j(2)} = v_{j(1)k}$ (i.e. the WCU b can board the journey $j(1)$ before the transfer point $w_{j(1)j(2)}$) and $s_b = v_{j(r)l}$, $a_b > a_{j(1)l}$, $l > k$ for $w_{j(r-1)j(r)} = v_{j(r)k}$ (i.e. the WCU b can get off the journey $j(r)$ after the transfer point $w_{j(1)j(2)}$).

The basic problem (abb. BP) is to find a set $\tilde{J}_F$ of daily duties for the set $F \subset F_0$ such that each WCU $b \in B$ is covered by a WCU($w_{j(1)j(2)}$, …, $w_{j(r-1)j(r)}$)-passable sequence $\vec{j} = j(1), \ldots, j(r) \in \tilde{J}_F$ and the cardinality $|\tilde{J}_F|$ is minimal, or, if such an "all WCU's covering" set does not exist, then to find such a set $\tilde{J}_F$ which maximizes the number of WCU's $b \in B$ covered by a WCU($w_{j(1)j(2)}$, …, $w_{j(r-1)j(r)}$)-passable sequence $\vec{j} = j(1), \ldots, j(r) \in \tilde{J}_{F0}$.

2 Solution of BP

Solvability of the problem BP will be examined first for a no transfer case

2.1 No Transfer Case

A WCU $b \in B$ is called direct traveler if there is at least one journey $j \in J$ meeting the following constraints:

C1: There exist integers k, l such that $o_b = v_{jk}$, $s_b = v_{jl}$ and $k < l$.

C2: If $a_b = 1440$ then $t_b \le t_{ik}$ else $a_b \ge a_{il}$.

If both constraints are valid for the journey j then one says that the trip of the WCU b is covered by the journey j and writes $j \blacktriangleleft b$.

If all WCU's are direct travelers then the problem BP is obviously NP-hard since its relaxed form after abandoning the daily duties formation represents a variant of the set covering problem (the trips of WCU's are covered by the set J) (see e.g. [3]). Therefore, looking for heuristic methods of solution is more hopeful than for the exact ones.

The next heuristic methods will not take into account the initial positions g_f of the WFV's $f \in F$. It is supposed, on the contrary, that the positions will be determined at the beginning of the resulting daily duties.

Heuristics Based on Set Covering and Consecutive Scheduling (HS)

The heuristics consists of the following steps:

1st step (initial)**:** The set $J_b = \{j \in J: j \blacktriangleleft b\}$ is found for each $b \in B$ and $J_B = \bigcup_{b \in B} J_b$.

The number of elements in the set J_B is denoted $|J_B|$ and $n = 0$ is put.

2nd step (initial)**:** The "classic" set covering problem is solved (e.g. by linear programming or by approximation algorithm [3]) for the set B covered by the set J_B. The resulting minimal covering subset of J_B is denoted J_{Bn} (there J_{B0}).

3rd step (recursive)**:** The minimal set of daily duties $\tilde{J}_n$ containing all journeys from the set J_{Bn} is created, e.g. by the method KASTOR [4], which is in the Czech and Slovak Republics successfully applied in the last two decades.

4th step (recursive)**:** If $|\tilde{J}_n| \leq |F|$ then the daily duties from the set $\tilde{J}_n$ are assigned to WFV's from the set F and the solution is over. If $|\tilde{J}_0| > |F|$ then $n = n + 1$ is put and the 5th step is to be done.

5th step: The maximum coverage problem (see e.g. [3]) is solved for the set B covered by the set J_B with the constraint $|J_{Bn}| = |J_{B(n-1)}| - 1$ for the cardinality of the resulting set J_{Bn}. Then the 3rd step is to be done again.

Heuristics of "Greedy" type (HG)

1st step (initial)**:** It is defined $\boldsymbol{J} = \varnothing$, $B'' = B$. The set $J_b = \{j \in J: j \blacktriangleleft b\}$ is found for each $b \in B''$ and $J_{B''} = \bigcup_{b \in B''} J_b$.

2nd step (recursive)**:** For each $j \in J_{B''}$ it is defined the set $B_j = \{b \in B'': j \blacktriangleleft b\}$. Let $|Bj|$ be the number of elements in the set B_j. Let $j' \in J_{B''}$ be the element with the maximal value $|Bj|$. Then it is defined $\tilde{j} = j'$ as a daily duty containing only one journey.

3rd step (recursive)**:** Let $\tilde{J}^*_{\tilde{j}}$ be the set of all simple (right and left) extensions of the daily duty $\tilde{j}$. If $\tilde{J}^*_{\tilde{j}} = \varnothing$ then the 4th step is to be done. Otherwise let $\tilde{j}' \in \tilde{J}^*_{\tilde{j}}$ be the daily duty maximizing the number $|B_{\tilde{j}'}|$, i.e. the number of WCU's $b \in B''$ such that $\tilde{j} \blacktriangleleft b$. Then it is defined $\tilde{j} = \tilde{j}'$ and the 3rd step is to be done.

4th step (recursive): It is defined $B' = B'' - \{b \in B'': \tilde{j} \blacktriangleleft b\}$, $J' = J \cup \{\tilde{j}\}$ and $J = J'$. If $B' = \varnothing$ or $|J| = |F|$ then the work is over and he set $\tilde{J}_F = J$ represents the solution of the basic problem. If $B' \neq \varnothing$ and $|J| < |F|$, then it is defined $B'' = B'$ and the 2nd step is to be done.

2.2 Transfer Case

In contrast to the previous case 2.1, it will now be considered that the WCU's can travel only with transfers i.e. his/her travel demand can be covered by a WCU($w_{j(1)j(2)}$, …, $w_{j(r-1)j(r)}$)-passable sequence of journeys $\vec{j} = j(1)$, …, $j(r)$. It is obvious, that it would be hardly possible to reformulate this version of BP into some kind of set covering problem. It is enough to imagine e.g. that a daily duty $\tilde{j}_1 = j_1(1),..., j_1(r_1)$ covers first part $j(1)$ of the sequence $\vec{j} = j(1)$, …, $j(r)$ of and the duty $\tilde{j}_2 = j_2(1),..., j_2(r_2)$ covers the second part $j(2)$. Then neither $\tilde{j}_1$

nor $\tilde{j}_2$ covers the trip of WCU but the union of them does. Therefore an extension of the heuristics of the type HS to the transfer case cannot be expected.

The case of greedy heuristics HG is much more hopeful. For the use in the transfer case, it will do a small reformulation of the 3rd step as follows. The set $B_{\tilde{j}'}$ will be defined as the set of journeys $b \in B''$ such that the journeys from the set $J \cup \{\tilde{j}\}$ cover one of possible trips of the WCU b.

3 Conclusion

The paper presents an exact formulation of the problem of covering the public transport demand of wheelchair users (WCU) in the case when wheelchair user friendly vehicles represent only a minor part of rolling stock. It is shown that it is an NP-hard problem. Further two heuristics are proposed for the solution of the particular case when all WCU's demand a direct travel without transfers. The first heuristics is based on set covering and bus scheduling techniques. The second one is of the greedy type. Finally it is shown that the slightly modified second heuristics can be used for the transfer case as well.

The future research could be focused on the use of mixed linear programming.

Acknowledgements

The authors would like to thank the Czech Science Foundation for the support of the project 402/12/2147 whose results are described in this paper.

References

[1] Belcher, M.J.H. and Frank, A.O.: Survey of the Use of Transport by Recipients of a Regional Electric Indoor/Outdoor Powered (EPIOC) Wheelchair Service. *Wheelchair Service*, 26 (2004), 563-575.

[2] De Winne, E.: Towards an integral accessible public area in the city. In: *The Sustainable City IV*: *Urban Regeneration and Sustainability* (Mander, U., Brebbia, C.A. and Tiezzi, E. eds.). Book Series: WIT Transactions on Ecology and the Environment. 93 (2006), 807-815.

[3] Hochbaum, D.C.: Approximating covering and packing problems: set cover, vertex cover, independent set, and related problems. In: *Approximation algorithms for NP-hard problems* (D.S. Hochbaum ed.), Ch. 3. PWS Publishing Co. Boston, MA, USA (1997), 94-143.

[4] Palúch, S.: The KASTOR System of Vehicle Scheduling Optimization (in Slovak). *Zborník VÚD Žilina*, No. 54, (1988), 249-257.

[5] Pressl, B., Mader, C. and Wieser, M.: User-Specific Web-Based Route Planning. In: *12th International Conference ICCHP 2010*, Vienna, Austria, July 14-16, (2010), 280-287.

Modification of the Method PRIVOL for Optimal Routing and Frequencing

Anna Černá[1], Jan Černý[2], Štefan Peško[3]

Abstract. PRIVOL is one of the methods that are successfully used to optimize the system of public transport lines in the Czech and Slovak Republic in the last decades. The main purpose of this paper is to describe some modifications of this method that allow to include in it some problems that have been solved "by hand" till this time.
It mainly concerns the following questions:
Q1: How to assign the vehicles to the selected routes in the case of a fleet which is heterogeneous in terms of vehicle size?
Q2: What to do if the demand is variable (in time) or elastic (with respect to the supplied frequency of service)?
The paper answers the questions in exact and/or in approximate form.

Keywords: routing, frequencing, public transport, optimization, PRIVOL, method.

JEL Classification: C61, C65, D85, O18
AMS Classification: 05C22, 05C35, 90B06

1 Introduction

This paper is focused on the problem of routing and frequencing in public transport, i.e. the optimization of the system of public transport lines. This topic was intensively studied during last two decades. The survey papers [12], [14] and [16] cite 20, 60 and 70 papers. One can see different methodological approaches. E.g. [1] focuses on the application of genetic algorithm, [6] describes a three-phases heuristics, which was successfully used in Rome, [7] connects mathematical programming approaches with decision-making and [4] describes a column generation approach. A man-machine approach, used e.g. in Sweden, is outlined by [2]. As one can see from [5], the issue was studied even in early seventies. The Czech contribution represents the paper [15].

It deals with the method PRIVOL (see e.g. [3]), which is successfully used in the Czech and Slovak Republic for many years. The goal of this paper is to extend the spectrum of particular cases which are solvable by PRIVOL.

PRIVOL arose in 80's as a reaction of Czechoslovak scientist to the paper [13] where the following approach was described: First, a set R_0 of "candidate" routes is created, manually by a transport engineer or with help of computer. Second, the configuration $R \subset R_0$ (to be operated) is selected by means of a non-linear programming minimizing the time losses of all passengers.

In that time, the available computers of Czechoslovak engineers were not suitable for such a method. The admissible size $|R|$ of the set R was about 10 which was not sufficient for practical purposes. Therefore, an attempt to abandon the non-linear objective function from [13] was made. The main problem was to find a new indicator which can replace passenger time losses but to converge to the same target. Finally, it was chosen the indicator y of the number of seats over the number of demanding passengers. It was used in two ways. The first, proposed in [8] and [9], took the new indicator in a constraint and minimized the necessary number SUM(x_r) of vehicles, where x_r is the number of vehicles assigned to the route $r \in R_0$. Of course, $x_r = 0$ means that r is not chosen into R. The second approach [10] left the number of vehicles in the constraint and to maximized new indicator. The latter approach is regarded as the basis of the method PRIVOL.

2 Basic Form of the Method PRIVOL

Let the network be given in the form of undirected graph G = (V, E) where V is the set of vertices (= nodes) and E is the set of edges (= segments). Let f_e be the demanded number of places for passengers passing through the edge $e \in E$ in one direction. Let the candidate set R_0 be given by means of the set E_r of all edges belonging to the route $r \in R_0$ and the total rough round running time t_r. Let n be the number of available vehicles, each of them

[1] Vysoká škola ekonomická, FM, Jarošovská 1117, Jindřichův Hradec, Czech Republic, cerna@fm.vse.cz.
[2] Vysoká škola ekonomická, FM, Jarošovská 1117, Jindřichův Hradec, Czech Republic, cerny@fm.vse.cz.
[3] Žilinská univerzita, FRaI, Univerzitná 8215/1, Žilina, Slovakia, pesko@ frcatel.fri.uniza.sk.

having the capacity c (= number of places for passengers). The problem is to find a positive real number y and a nonnegative integer x_r for each $r \in R_0$ such that

$$y \to \max \tag{1}$$

$$\sum_{r \in R_0} x_r = n \tag{2}$$

$$\sum_{r: e \in E_r} \frac{60c}{t_r} x_r \geq f_e y \quad \text{for each } e \in E \tag{3}$$

2.1 Illustrative Example

Suppose that a town has only one route of urban transport connecting four vertices $u - v - w - x$ and several intermediate stops, not important in this case. Suppose that the running times of buses through the edges (u, v), (v, w) and (w, x) are equal to 9 minutes. Suppose that in the morning peak there are the following important flows of passengers (the others are negligible): from u to v the flow q_{uv} = 50 passengers per hour, further q_{uw} = 400, q_{ux} = 100, q_{vw} = 250, q_{vx} = 450 and q_{wx} = 20. Therefore the cumulated passenger demand f_{uv} = 550, f_{vw} = 1200 and f_{wx} = 570. Suppose further that the transport company uses 12 buses with capacity c = 90 passengers, having the total round running time from u to x and back (including stays at terminals) 60 min and, therefore, the headway 5 min and the total supply of 1080 places for passengers per hour, much less than the demand 1200.

Of course, the traveling public protests against this situation. They ask to add at least 3 buses, but without success.

In this situation, the candidate set R_0 may be chosen e.g. in the following way: $R_0 = \{r_1, r_2, r_3\}$ where

r_1 is the original route u-v-w-x with $t_1 = 60$,

r_2 is a new possible route u-v-w with $t_2 = 40$,

r_3 is a new possible route v-w-x with $t_3 = 40$.

Then the constraints (1), (2), (3) turn to the form

$$y \to \max \tag{1*}$$

$$x_1 + x_2 + x_3 = 12 \tag{2*}$$

$$\begin{aligned} 90x_1 + 135x_2 &\geq 550y \\ 90x_1 + 135x_2 + 135x_3 &\geq 1200y \\ 135x_2 + 135x_3 &\geq 570y \end{aligned} \tag{3*}$$

The (optimal) integer solution is $x_1 = 0$, $x_2 = 6$, $x_3 = 6$ and $y = 1.35$, i.e., the route r_1 should be abandoned and the fleet should be assigned half and half to the two new lines. Then the traveling becomes comfortable: On the edge (u, v) the demand is 550, the supply is 810, i.e. 47,3% more, on (v, w) it is 1200 vs. 1620, i.e. 35% more and on (w, x) it is 810 vs. 570, i.e. 42,1% more.

3 Known Modifications

In this chapter, the already (for many years) used modifications are presented. In addition to the denotations introduced in the previous chapters, the following ones will be used in the sequel:

f_{vw} = number of passengers from the vertex v to the vertex w per hour,
n_k = total number of vehicles in the k^{th} system,
c_k = capacity of the vehicle in the k^{th} system.

3.1 Fleet Minimization

This is the "classic" formulation from [9] which minimizes the rolling stock provided that the supplied capacity is anywhere at least equal to the demand:

Let the graph G and the set R_0 be given. The problem is to find a nonnegative integer x_r for each $r \in R_0$ such that

$$z = \sum_{r \in R_0} x_r \to \min \tag{4}$$

$$\sum_{r:e \in E_r} \frac{60c}{t_r} x_r \geq f_e \quad \text{for each } e \in E \tag{5}$$

Remark. If this model is applied in the case of 2.1, the solution will be $x_1 = 0$, $x_2 = 5$, $x_3 = 5$ with the supply demand ratios 675/550 = 1,227 (+22,7%) on (u, w), 1350/1200 = 1,125 (+12,5%) on (v, w) and 675/570 = 1,184 (+18,4%).

3.2 Limitation of Transfers

After having solved the problem from the chapter 2 or 3.1, it may happen that the resulting percentage of passengers having at least one transfer is not acceptable. Although this problem is not explicitly described in literature, a natural way of solution is commonly used in practice. It is necessary to find such pair of vertices v, w, that no route in R contains both vertices v and w and the flow f_{vw} between them is maximal. Afterwards the set $R_{vw} = \{r \in R_0: r \text{ contains both vertices } v, w\}$ is observed and if $R_{vw} = \varnothing$ then a new route connecting v and w is added into it. Finally, the following constraint is added to the problem:

$$\sum_{r \in R_{vw}} x_r \geq 1 \tag{6}$$

and the problem is solved again. Then, at least one route from the new resulting set R contains both v and w. I.e. a new possibility of direct travel from the vertex v to w arose for f_{vw} passengers per hour. Therefore,, the number of directly traveling passengers increases. If it is not yet satisfactory, another constraint of the type (6) can be added etc.

3.3 Two or Three Different Subsystems

It happens that two or three modes (e.g., tram, trolleybus and bus), are operating on surface level in a town. Then (see [11]] the set R_0 of candidate routes is partitioned into two or three disjoint subsets R_{01} (candidate routes for the first subsystem), R_{02} (candidates for the second subsystem) etc. Then the constraint (2) is modified to the constraint

$$\sum_{r \in R_{0k}} x_r = n_k \quad \text{for } k = 1, 2 \text{ (and 3 if necessary)} \tag{7}$$

where n_1, n_2, (n_3) means the available number of vehicles of the 1st, 2nd (3rd) subsystems respectively and the constraint (3) is modified to the constraint

$$\sum_{r:e \in E_r} \frac{60c_k}{t_r} x_r \geq f_e y \quad \text{for each } e \in E \text{ and } k = 1,2 \text{ (or 3 if necessary)} \tag{8}$$

3.4 Abandoning the Integer Variables

It happens that it is not known what rolling stock will serve the routes sought in the chapter 2. Then, as shown in [10], the meaning of the route selecting variable x_r should be turned into the number of passenger places in the vehicles and the total number of vehicles n should be turned into the expected total number of places for passengers in the rolling stock. Then the constraints (1) and (2) remain unchanged, while the constraint (3) turns into

$$\sum_{r:e \in E_r} \frac{60}{t_r} x_r \geq f_e y \quad \text{for each } e \in E \tag{9}$$

and a further constraint is added:

$$x_r \geq c_0 \quad \text{for each } r \in R_0 \tag{10}$$

where c_0 denotes the minimum possible vehicle capacity (e.g., c_0 =28). This avoids unfeasible results as x_r = 5.3.

4 New Modifications

Now, five not yet known modifications of the method PRIVOL, are to be presented. They are motivated by needs of urban transport practice.

The first modification deals with a rolling stock which is heterogeneous in terms of the size (i.e. capacity) of vehicles.

4.1 Heterogeneous Rolling Stock but Homogeneous Groups of Vehicles on Routes

Suppose that there are k = 2 or 3 types of vehicles with different sizes, i.e. with different capacities. It is a cascade type problem. The first step of the cascade assigns the type k of vehicle to each route r following e,g, some transport engineering aspects (necessity of a small vehicle because of sharp turns in the route etc.). Then the problems turns into the one presented in 3.3.

4.2 Heterogeneous Rolling Stock and Heterogeneous Groups of Vehicles on Routes

Suppose again that there are k = 2 or 3 types of vehicles with different sizes but the partition of the set R_0 mentioned in 4.1 is not possible. Then it is again the problem of cascade type, but other than in the previous case. The first step resolves the problem 3.4 with n expressing the total capacity of the rolling stock. E.g., if k = 3 and the number of the k^{th} type of vehicles is n_k and their capacity is c_k then $n = c_1n_1 + c_2n_2 + c_3n_3$.

Denote the resulting values $x_r = a_r$ for each $r \in R_0$ and define the variable z_{rk} as number of vehicles of the k^{th} type assigned to the route r for each $r \in R_0$. Then the second step of the cascade is to find nonnegative integers z_{rk}, $r \in R_0$, k = 1, 2, 3 and a real number y' such that

$$\sum_{r \in R_0} z_{rk} = n_k \quad \text{for each } k = 1, 2, 3 \tag{11}$$

$$\sum_{k=1,2,3} c_k z_{rk} \geq a_r y' \quad \text{for each } r \in R_0 \tag{12}$$

$$y' \rightarrow \max \tag{13}$$

Remark 1. The number a_r represents "ideal" capacity of the router r, the sum $z_{r1} + z_{r2} + z_{r3}$ represents the final result prepared for implementation to the service. The constraint (12) ensures that the relative satisfaction of the route r is at least y' and (13) ensures that it is the greatest possible.

4.3 Limited Number of Routes

In the Czech public transport community is widespread unwritten principle that the number of urban routes should not exceed some number related to the population p of the town. The authors feel that it may be something like $0.2p$ if p is expressed in thousands. E.g., if the population is 25 thousands, the number of routes should not exceed 5. Hence, it may happen that the city authorities may ask the transport engineers to keep the number not exceeding the given limit l^*.

If the basic form of PRIVOL (Chapt. 2) is to be used than it is necessary to introduce new binary variables q_r for $r \in R$, where q_r = 1 if r is chosen to R and q_r = 0 otherwise. Then the following constraints should be added to (1), (2) and (3):

$$10^{-6}x_r \leq q_r \leq x_r \leq 10^6 q_r \quad \text{for each } r \in R_0 \tag{14}$$

$$\sum_{r \in R_0} q_r \leq l^* \tag{15}$$

4.4 Limited Number of Routes in Several Time Periods

Suppose that the route set optimization should be found by basic PRIVOL for m different time periods, e.g., morning peak, morning saddle, afternoon peak, evening, weekends etc. It happens that the optimal set of routes R_j for the j^{th} period completely differs from the others. The passengers do not like "the same routes" (more precisely, routes having the same number) passing different places in different time periods. On the other hand, there are reasons for keeping the total number of routes under a limit l^*. Of course, it is acceptable to have different headways i.e. different frequencies of the same routes in different time periods. Then the following problem could be formulated: to find a positive real number y and a nonnegative integers q_r, x_{rj} for each $j = 1, \ldots, m$ and $r \in R_0$ such that

$$y \to \max \tag{16}$$

$$\sum_{r \in R_0} x_{rj} = n_{oj} \quad \text{for } j = 1, \ldots, m \tag{17}$$

$$\sum_{r:e \in E_r} \frac{60c}{t_r} x_{rj} \geq f_{ej} y \quad \text{for and each } e \in E \text{ and } j = 1,\ldots,m \tag{18}$$

$$10^{-6} x_{rj} \leq q_r \leq x_{rj} \leq 10^6 q_r \quad \text{for each } r \in R_0 \text{ and } j = 1, \ldots, m \tag{19}$$

$$\sum_{r \in R_0} q_r \leq l^* \tag{20}\equiv(14)$$

where n_{oj} is the number of vehicles in service in the j^{th} time period and f_{ej} be the demanded number of places for passengers passing through the edge $e \in E$ in one direction in the j^{th} time period.

4.5 Elastic Demand with Respect to Supply

The title of the paragraph represents a complex problem even for the basic PRIVOL, since the elasticity could be different at the different place and time. In the sequel, a simplified problem and model will be presented for the case of 3.1. Suppose that on each edge the demand $f_e = b + ds_e$ where s_e represents the supplied number of places for passengers passing through the edge $e \in E$ in one direction and b is a positive real number, $b < 1/f_e$ for each $e \in E$. Then the constraint (5) turns to

$$\sum_{r:e \in E_r} \frac{60c}{t_r} x_r \geq f_e (1 + b \sum_{r:e \in E_r} \frac{60c}{t_r} x_r) \Leftrightarrow (1 - bf_e) \sum_{r:e \in E_r} \frac{60c}{t_r} x_r \geq f_e \text{ for each } e \in E \tag{21}$$

5 Conclusions

In the paper, it was shown that the method PRIVOL is flexible in the sense of its adaptability to different practical situations as

- Heterogeneity of rolling stock;
- different demand in different time period;
- elasticity of demand with respect to the supply.

The authors are convinced that that is not the end, that there are several other modifications, e.g. towards a limit of the total length of the subnet used by the selected set R.

Acknowledgements

The authors would like to thank the Czech Science Foundation for the support of the project 402/12/2147 whose results are described in this paper.

References

[1] Agrawal, J. and Mathew, T. V.: Transit route design using parallel genetic algorithm. *J. Comput. Civ. Eng.*, **18** (2004), 248–256.

[2] Andreasson, I.: A method for the analysis of transit networks. Roubens, M. (Ed.): *Advances of operations research*, North Holland, Amsterdam, 1977, p. 1–8.

[3] Blatoň, M.: Komparace vybraných metod zabývajících se návrhem sítě linek MHD (Comparison of selected methods for urban transport routing – in Czech). *Perner Contact* **4**(2009), No. III, 6-14.
[4] Borndörfer, R., Grötschel, M. and Pfetsch, M.E.: A column generation approach to line planning in public transport. *Transp. Sci*. **41** (2007), 123–132.
[5] Byrne, B. and Vuchic, V.: Public transportation line positions and headways for minimum cost. F. G. Newell, F. G. (Ed.): *Traffic flow and transportation*, Elsevier, New York, 1972, p. 347–360.
[6] Carrese, S., and Gori, S.: An urban bus network design procedure. M. Patriksson and M. Labbé (Eds.), *Transportation planning*, Kluwer Academic Publisher, Dordrecht, The Netherlands, 2002 p. 177–195.
[7] Ceder, A.: Operational objective functions in designing public transport routes. *J. Adv. Transp.* **35** (2001), 125–144.
[8] Černý, J.: Poznámka k jednotným intervalom v systémoch MHD (Note to unified intervals in urban transport systems – in Slovak). *Zborník prác výskumného ústavu dopravného* No. 39, Žilina, Slovakia, 1983, p. 157-175.
[9] Černý, J.: Metódy konštrukcie cestovných poriadkov osobnej dopravy (Scheduling methods in passenger transport – in Slovak). Skýva L. (Ed.) *Optimalizačné metódy v riadení dopravy*. Dom techniky a VŠDS Žilina, 25-29.6.1984 Púchov, p. 74-128.
[10] Černý, J. and Černá, A.: Linear optimization of urban bus routes and frequencies. *Czechoslovak Journal for Operations Research* **1**(1992), 207-217.
[11] Černý, J. and Černá, A.: Erlander Principle in Managerial Decision Making on Czech and Slovak Urban Transport Routes. *E&M Economics and Management* **16**(2013), 93-100.
[12] Desaulniers, G. and Hickman, M. D. (2007): Public Transit. In C. Barnhart and G. Laporte (Eds.): *Handbook in OR & MS,transportation*, Vol. 14, Elsevier, Amsterdam, 2007, p. 69-127.
[13] Erlander, S. and Schéele, S.:A mathematical programming model for bus traffic in a network. In D.J. Buckley (Ed.): *Transportation and Traffic Theory*. Elsevier, New York, 1974, p. 581-605.
[14] Kepaptsoglou, K. and Karlaftis, M. G.: Transit Route Network Design Problem: Review. *Journal of Transportation Engineering*, **135** (2009) 491-505.
[15] Přibyl, V.: Solution of the bus route design problem. *Communications*, **11** (2009), 25-28.
[16] Schöbel, A.: Line planning in public transportation: models and methods. *OR Spectrum*, **34** (2012), 491–510.

Interval data and sample variance: Computational issues

Michal Černý[1]

Abstract. We study the problem of computation of bounds of sample variance over an interval dataset. The tight lower bound can be computed in polynomial time by reduction to convex quadratic programming. On the other hand, computation of the upper bound is an NP-hard problem. We inspect special cases which admit polynomial-time computation and design a simple heuristic algorithm. We support the study by a computational experiment.

Keywords: interval data; sample variance; computational complexity

JEL classification: C61
AMS classification: 90C60

1 Introduction: Problem formulation

This article is a contribution to the following general question. Let $x_1, \ldots, x_n$ be a collection of one-dimensional data which are unobservable. What is observable is a collection of intervals

$$\boldsymbol{x}_1 := [\underline{x}_1, \overline{x}_1], \ \ldots, \ \boldsymbol{x}_n := [\underline{x}_n, \overline{x}_n] \tag{1}$$

such that it is guaranteed that

$$x_i \in \boldsymbol{x}_i \quad \text{for all} \ \ i = 1, \ldots, n. \tag{2}$$

We study the general situation when we cannot make any further assumptions than (2) on the distribution of $x = (x_1, \ldots, x_n)^{\mathrm{T}}$ given $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^{\mathrm{T}}$.

Let a statistic $S = S(x_1, \ldots, x_n)$ be given. Having no information on the distribution of x given $\boldsymbol{x}$, the only information about S we can infer from the observable data $\boldsymbol{x}$ is the pair of values

$$\underline{S} = \underline{S}(\boldsymbol{x}) = \inf\{S(x_1, \ldots, x_n) : \ x_i \in \boldsymbol{x}_i, \ i = 1, \ldots, n\}, \tag{3}$$

$$\overline{S} = \overline{S}(\boldsymbol{x}) = \sup\{S(x_1, \ldots, x_n) : \ x_i \in \boldsymbol{x}_i, \ i = 1, \ldots, n\}. \tag{4}$$

The interval $[\underline{S}, \overline{S}]$ bounds S (given $\boldsymbol{x}$) a.s. and provides us with at least partial information about S.

The interval $[\underline{S}, \overline{S}]$ has been studied for a class of usual statistics, such as sample mean, sample variance, t-ratio, sample entropy, sample correlation and many others. In general, the problem of observability of $\boldsymbol{x}$ instead of x is sometimes referred to as *incompleteness of data* or a particular form of *partial identification*. This problem arises often in practice: for example, data could have been rounded, categorized or censored, or they can be simply affected by measurement error such that we cannot make further assumptions on its distribution.

In this text we deal with sample variance

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \widehat{\mu})^2,$$

where $\widehat{\mu} = \frac{1}{n} \sum_{j=1}^{n} x_j$. We will study algorithms for computation of the interval $[\underline{\widehat{\sigma}^2}, \overline{\widehat{\sigma}^2}]$.

[1]University of Economics Prague, Department of Econometrics, Winston Churchill Square 4, 13067 Prague, Czech Republic, cernym@vse.cz

2 Why a heuristic method?

Many problems in statistics can be formulated as optimization problems; two well-known examples are minimization of sum of squares and maximization of the likelihood function. Problems (3) and (4) are of optimization nature, too. Dealing with sample variance, problems (3) and (4) take the form

$$\underline{\widehat{\sigma}^2} = \underline{\widehat{\sigma}^2}(\boldsymbol{x}) = \min \left\{ \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right)^2 \,:\, x_1 \in \boldsymbol{x}_1,\ \ldots,\ x_n \in \boldsymbol{x}_n \right\}, \tag{5}$$

$$\overline{\widehat{\sigma^2}} = \overline{\widehat{\sigma^2}}(\boldsymbol{x}) = \max \left\{ \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right)^2 \,:\, x_1 \in \boldsymbol{x}_1,\ \ldots,\ x_n \in \boldsymbol{x}_n \right\}. \tag{6}$$

2.1 Complexity of computation of $\underline{\widehat{\sigma}^2}$ and $\overline{\widehat{\sigma^2}}$

Observe that (5) is a convex optimization problem. Thus it is not difficult to prove:

Theorem 1 (Ferson et al. [4]; Černý and Hladík [3]). *There exists a polynomial-time algorithm for computation of* $\underline{\widehat{\sigma}^2}$. □

So the value $\underline{\widehat{\sigma}^2}$ can be computed exactly and efficiently and we do not need any heuristic algorithms.

Remark. The paper [4] is not interesting only for the observation stated in Theorem 1; it also introduces a remarkable *quadratic time* algorithm for computation of $\underline{\widehat{\sigma}^2}$ which does *not* require special optimization techniques (such as the ellipsoid algorithm or interior-point methods).

Unfortunately, the problem (6) is not convex. The trouble is indeed serious:

Theorem 2 (Ferson et al. [4]). *Computation of* $\overline{\widehat{\sigma^2}}$ *is an* ***NP****-hard problem.* □

And the bad news goes further:

Theorem 3 (Ferson et al. [4]; Černý and Hladík [3]). *Approximate computation of* $\overline{\widehat{\sigma^2}}$ *with an arbitrary absolute error is an* ***NP****-hard problem.* □

Theorem 3 should be understood as follows: should we admit *any* (possibly large) fixed error $\Delta > 0$, if there exists a polynomial-time algorithm $A_\Delta(\boldsymbol{x})$ such that $|A_\Delta(\boldsymbol{x}) - \overline{\widehat{\sigma^2}}| \leq \Delta$, then $\boldsymbol{P} = \boldsymbol{NP}$.

3 Exact exponential-time algorithm for $\overline{\widehat{\sigma^2}}$

Theorem 2 shows that we can expect only algorithms with time 2^n or worse. Knowing this, it is good news that the "or worse" case can be excluded. Said otherwise, the value $\overline{\widehat{\sigma^2}}$ can be indeed computed in time 2^n. This fact follows from the following lemma.

Lemma 4 (proof omitted). There exist $x_i^* \in \{\underline{x}_i, \overline{x}_i\}$, $i = 1, \ldots, n$, such that $\overline{\widehat{\sigma^2}} = \sigma^2(x_1^*, \ldots, x_n^*)$. □

It follows that computation of $\overline{\widehat{\sigma^2}}$ can be reduced to 0-1 programming with n variables:

$$\max_{\xi_1,\ldots,\xi_n} \{\sigma^2(\underline{x}_1 + \xi_1(\overline{x}_1 - \underline{x}_1), \ldots, \underline{x}_n + \xi_n(\overline{x}_n - \underline{x}_n)) :\ \xi_i \in \{0,1\},\ i = 1, \ldots, n\}. \tag{7}$$

Or, said otherwise, it suffices to inspect all 2^n vertices of the cube $\boldsymbol{x}_1 \times \cdots \times \boldsymbol{x}_n$.

The negative complexity-theoretic results of Theorems 2 and 3 justify that usage of heuristic algorithms for $\overline{\widehat{\sigma^2}}$ will be necessary, especially when processing larger datasets.

4 Polynomial-time special cases

Before turning to heuristics, it is always necessary to inspect special cases which are solvable in polynomial time. So far there are (at least) two known such special cases, covering many datasets appearing in practice (but still, many datasets remain uncovered).

4.1 The first special case

The first special case is based on the following result.

Theorem 5 (Ferson et al. [4]). *Let $x_i^\Delta := \frac{1}{2}(\overline{x}_i - \underline{x}_i)$ be the radius of $\boldsymbol{x}_i$ and let $x_i^c := \frac{1}{2}(\overline{x}_i + \underline{x}_i)$ be the center of $\boldsymbol{x}_i$. If*

$$[x_i^c - \tfrac{1}{n}x_i^\Delta, x_i^c + \tfrac{1}{n}x_i^\Delta] \cap [x_j^c - \tfrac{1}{n}x_j^\Delta, x_j^c + \tfrac{1}{n}x_j^\Delta] = \emptyset, \quad 1 \le i < j \le n, \tag{8}$$

then $\overline{\widehat{\sigma^2}}$ can be computed in quadratic time. □

Observe that the condition (8) is efficiently testable. Being strict, we can say that *every* algorithm which is claimed to be "good" *must* start with the test (8) and if the test is successful, the algorithm simply switches to the exact quadratic-time method.

4.2 The second special case

The second special case is based on the following result.

Theorem 6 (Černý and Hladík [3]). *Let the values $\underline{x}_1, \overline{x}_1; \ldots; \underline{x}_n, \overline{x}_n$ be integer. There exist an algorithm for computation of $\overline{\widehat{\sigma^2}}$ working in time $p(n, M)$, where p is a polynomial and*

$$M = \max\{|\underline{x}_1|, |\overline{x}_1|, \ldots, |\underline{x}_n|, |\overline{x}_n|\}. \quad \square$$

The algorithm of Theorem 6 is not a polynomial-time algorithm in general, but it is polynomial e.g. when M is bounded from above by a constant. So this special case is useful when the numbers $\underline{x}_1, \overline{x}_1; \ldots; \underline{x}_n, \overline{x}_n$ are "not too large". In that case, we can efficiently compute $\overline{\widehat{\sigma^2}}$ even for large n.

Of course, the integrality assumption can be removed. Say for example that the numbers $\underline{x}_1, \overline{x}_1; \ldots; \underline{x}_n, \overline{x}_n$ are expressed as decimal numbers with one decimal place; then a premultiplication by the factor 10 converts the data into integers (at the cost of an increase in M).

The nice property of Theorem 6 is that $T := p(n, M)$ can be computed in advance. If the required computation time T is acceptable, then it is appropriate to prefer the algorithm of Theorem 6 to heuristics.

Remark. Preprocessing of data might be useful. A shift of data can reduce the value M. Defining

$$\underline{M} = \min_{i=1,\ldots,n} \underline{x}_i, \quad \overline{M} = \max_{i=1,\ldots,n} \overline{x}_i, \quad M^\Delta = \lfloor \tfrac{1}{2}(\overline{M} + \underline{M}) \rfloor,$$

we can replace the dataset $[\underline{x}_1, \overline{x}_1], \ldots, [\underline{x}_n, \overline{x}_n]$ by the shifted dataset

$$[\underline{x}_1 - M^\Delta, \overline{x}_1 - M^\Delta], \ \ldots, \ [\underline{x}_n - M^\Delta, \overline{x}_n - M^\Delta].$$

5 Two simple heuristics

5.1 The first heuristic

By Lemma 4 we know that the value $\overline{\widehat{\sigma^2}}$ is attained in a vertex of the box $\boldsymbol{x}_1 \times \cdots \times \boldsymbol{x}_n$. Let V be the set of these vertices. Heuristic methods often utilize various strategies for inspection of V, starting from an initial point. In general, it is very important to give a good starting point.

Let $A = I - \frac{1}{n}E$, where I is the unit matrix and E is the all-one matrix. Clearly $(n-1)\widehat{\sigma}^2(x) = x^{\mathrm{T}}Ax$. Consider the starting point

$$x^* = x^c + \mathrm{diag}(\mathrm{sgn}(Ax^c))x^\Delta \tag{9}$$

(this idea is by Hladík, unpublished so far). It simply tells that we look at the gradient of the quadratic form $x^{\mathrm{T}}Ax$ in the center point x^c and find a vertex in the "direction" of the gradient.

5.2 Computational experiments

Assume that interval data were generated by the process

$$\underline{x}_i = -10U_i, \qquad \overline{x}_i = 10V_i, \qquad i = 1, \ldots, n,$$

where U_i and V_i are independent $(0,1)$-uniformly distributed random variables. For small values of n (here, $n = 8, \ldots, 15$) we can use the exact 2^n-algorithm (7) to find the optimal vertex of the cube $\boldsymbol{x}_1 \times \cdots \times \boldsymbol{x}_n$ and thus the correct value $\overline{\widehat{\sigma^2}}$. The following table summarizes results of computational experiments with the heuristic (9). To explain the figures in the table, let us look at row $n = 8$. The value 0.983 means that on average we observed $\frac{\overline{\widehat{\sigma^2}}_{\text{HEURISTIC}}}{\overline{\widehat{\sigma^2}}_{\text{CORRECT}}} = 0.983$. In 55% of cases, the heuristic (9) found the correct vertex exactly. In the 45% cases, when the heuristic found an incorrect vertex, the average Hamming distance between the optimal vertex and the vertex found by the heuristic was 2.55.

	Average Hamming distance from the optimal vertex	% of simulations when the optimal vertex was found exactly	Relative distance to the optimal value $\overline{\widehat{\sigma^2}}$
$n = 8$	2.55	55%	0.983
$n = 9$	2.11	55%	0.990
$n = 10$	2.50	60%	0.989
$n = 11$	2.00	44%	0.985
$n = 12$	2.85	36%	0.978
$n = 13$	3.11	54%	0.991
$n = 14$	2.83	40%	0.984
$n = 15$	2.55	44%	0.988

5.3 Polynomial-time computable lower and upper bounds

In computational geometry, "rounding" of a complex geometric object G (e.g. a box or a polytope) means replacement of G by another object G' which is easy-to-handle and it approximates the object G in some well-defined sense (here it is not necessary to be more precise). In this section we introduce an approach motivated by the theory of Löwner-John ellipsoids; for details see e.g. [1].

Let $\boldsymbol{x}$ be given. As the first ingredient we consider the pair of *Löwner-John ellipsoids*

$$\underline{\mathcal{E}} = \{\xi \in \mathbb{R}^n : (\xi - x^c)^{\mathrm{T}} (\mathrm{diag}(x^\Delta))^{-2} (\xi - x^c) \leq 1\},$$
$$\overline{\mathcal{E}} = \{\xi \in \mathbb{R}^n : (\xi - x^c)^{\mathrm{T}} (\mathrm{diag}(x^\Delta))^{-2} (\xi - x^c) \leq n\}.$$

Clearly we have

$$\underline{\mathcal{E}} \subseteq \boldsymbol{x} \subseteq \overline{\mathcal{E}}; \tag{10}$$

moreover, the ellipsoids have the property that $\underline{\mathcal{E}}$ is the largest inscribed ellipsoid in $\boldsymbol{x}$ and $\overline{\mathcal{E}}$ is the smallest circumscribed ellipsoid of $\boldsymbol{x}$. The second ingredient is in the following result.

Theorem 7 (Ye [8]; Yıldırım [9])**.** *Let Q be any symmetric matrix and let $\mathcal{E}$ be an ellipsoid. The optimization problem $\max\{x^T Q x : x \in \mathcal{E}\}$ can be solved in polynomial time.* □

So we can efficiently solve both problems

$$\max \left\{ \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right)^2 : (x_1, \ldots, x_n)^{\mathrm{T}} \in \underline{\mathcal{E}} \right\},$$
$$\max \left\{ \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right)^2 : (x_1, \ldots, x_n)^{\mathrm{T}} \in \overline{\mathcal{E}} \right\}.$$

Let x^*, x^{**} denote their respective optimal solutions — by (10) we have the desired bounds

$$x^* \leq \overline{\widehat{\sigma}^2} \leq x^{**}.$$

Remark. The algorithm of Theorem 7 requires nontrivial optimization techniques; in particular, it is based on an interior point algorithm for positive semidefinite programming.

Remark. In this section we used *geometric rounding* in the following sense: we have "rounded" the box $\boldsymbol{x}$ to an ellipsoid $\underline{\mathcal{E}}$ (or $\overline{\mathcal{E}}$) which is "easier to work with" in the sense that maximization of a quadratic function over the rounded object can be done in polynomial time.

6 Conclusions

Though the problem of computation of $\overline{\widehat{\sigma}^2}$ is NP-hard, we have shown that the situation is not so bad. There exist special cases which are tractable in polynomial time. Moreover, we have designed two heuristic approaches. The first one was tested in a computational experiment; results are promising, since the heuristic seems to approximate the correct value $\overline{\widehat{\sigma}^2}$ at 98% (in one particular model). The second heuristic, based on a combination of Löwner-John ellipsoids and semidefinite programming, is a polynomial-time method giving both upper and lower bound on the value $\overline{\widehat{\sigma}^2}$.

Acknowledgment

The work was supported by the Czech Science Foundation under Grant No P403/12/1947.

References

[1] Černý, M.: Goffin's algorithm for zonotopes. *Kybernetika* **48**, 2012, 890–906.

[2] Černý, M., Antoch, J., and Hladík, M.: On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data, *Information Sciences* **244** (2013), 26–47.

[3] Černý, M., and Hladík, M.: Complexity of computation and approximation of the t-ratio over one-dimensional interval data, submitted (2014).

[4] Ferson, S., Ginzburg, L. Kreinovich, V., Longpré, L., and Aviles, M.: Exact bounds on finite populations of interval data, *Reliable Computing* **11** (2005), 207–233.

[5] Gladysz, B., and Kasperski, A.: Computing mean absolute deviation under uncertainty. *Applied Soft Computing* **10** (2010), 361–366.

[6] He L. T., and Hu C.: Impacts of interval computing on stock market variability forecasting. *Computational Economics* **33** (2009), 263–276.

[7] Kreinovich, V.: Why intervals? A simple limit theorem that is similar to limit theorems from statistics, *Reliable Computing* **1** (1995), 33–40.

[8] Ye., Y.: On affine scaling algorithms for nonconvex quadratic programming. *Mathematical Programming* **56** (1992), 285–300.

[9] Yıldırım, E. A.: On the minimum volume covering ellipsoid of ellipsoids, *SIAM Journal of Optimization* **17** (2006), 621–641.

The cost of acceptance sampling in case of alternative qualifications of products due to many features

Magdalena Chmielińska[1]

Abstract. The acceptance sampling is the settlement procedure based on a sample randomly drawn from a larger batch of quality in the controlled batch. Conducted in factories acceptance control must not only be effective but also must be justified from an economic point of view. The companies monitored so the costs associated with it.
Developed and described in the literature patterns enable the evaluation costs of operating the acceptance control, but only when a product is evaluated only with respect to one of its feature. In practice, however, often controlled by more than one characteristic of the product produced. This raises the need to estimate the costs of operating the acceptance control in just such a case. This problem seems to be particularly interesting in the case of qualifications errors.

Keywords: Acceptance sampling, Alternative classification of many variables, Quality costs.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Ensuring the appropriate level of quality of manufactured products is the purpose of statistical quality control. The use of appropriate method of quality control guarantees adjusting the level of quality of the products to the requirements of potential customers. However, use of these method requires some outlays. Not only the monitoring the level of quality of manufactured products but also the monitoring the cost of quality control is necessary to ensure that the appropriate sales market, because only products characterized by an appropriate price – quality ratio are desired by customers. Hence the fundamental problem in sampling inspection is to strike the appropriate balance between quality and costs.

This article attempts to develop formulas for the expected cost of acceptance quality control where a controlled product is assessed alternatively due to more than one feature. This look at the cost of quality control is innovative because described in the literature patterns relate only to the case when a product is evaluated only with respect to one of its feature. In this article it was also considered an incomplete diagnostic skills, manifested by the presence of qualifications errors. In both considered cases (complete and incomplete diagnostic skills) it were assumed that the defectiveness of product due to individual assessed features is the same.

2 Description of acceptance sampling plan

Acceptance sampling plan is the procedure of deciding the quality of batch on the basis of the sample, which is randomly selected from this batch. It helps determine which lots of items are acceptable and which are unacceptable and should be reworked and corrected. As a method of "off – line" statistical quality control, acceptance sampling plan applies at a time when production of the batch is completed. Acceptance testing are carried out at this moment because their purpose is to protect the recipient from taking products that do not meet its quality requirements.

An important measure of the performance of on acceptance sampling plan is the operating characteristic curve. It is called in short OC curve and it shows the probability of accepting the lot versus the lot fraction defective at the used acceptance sampling plan [7]. In a slightly different formulation the OC curve shows what will be the percentage of accepted batches, if it is shown to receive a very large number of parties with a certain fraction defective. When the batches fraction defective is equal zero, all batches are accepted whence $L(p)=1$. As defectiveness of batches increases, the probability of accepting the lot decreases, until it is zero at batches

[1] University of Economics in Katowice, Department of Statistic, ul. Bogucicka 14, 40-226 Katowice, Poland, chmielinska.magda@gmail.com

defectiveness is equal one. In [2] we can see, that the OC curve for the attributes acceptance sampling plan can be written using the following formula:

$$L(p) = P(Z \le c) = \sum_{k=0}^{c} \frac{\binom{N-D}{n-k}\binom{D}{k}}{\binom{N}{n}} \tag{1}$$

where:

Z – random variable denoting the number of defective items in the randomly selected sample,
c – acceptance number,
N – the lot size,
n – the sample size,
D – the number of defective items in the lot,
k – the number of defective items in the sample.

In economic practice, the produced element has often more than one feature that can and should be controlled, assessed. This raises, the need for taking into account the multiple alternative qualifications into procedure of acceptance inspection occurs.

3 The cost of quality control

The concept of quality control costs means all enterprise's costs connected with conduction quality control. Iwasiewicz [2] shows, that costs of prevention, costs of examination and estimation, loss of internal lack and loss of external lack were distinguished among costs of statistical quality control. Total quality control cost is a sum of all distinguished groups of costs.

The total acceptance control cost consists of costs of examination and estimation, loss because of internal and external lack. The costs of examination and estimation depend on single costs of control (k_k) and on amount of controlled pieces. To signify as N – size of batch, n – size of sample, D – amount of defective elements in batch, d – amount of defective elements in sample, control cost by application the single – stage acceptance sampling, as Kończak [5] shows, are following:

$$K_k(p) = L(p)nk_k + (1 - L(p))Nk_k \tag{2}$$

In [5] we can see also, that loss of internal lack depends on cost of exchange (or repair) of defective element (k_{bw}) and on amount of damage elements detected in sample, in case of acceptance batch, or on amount of damage elements detected in batch, in case of its rejection. It can be written as:

$$K_{bw}(p) = L(p)\sum_{d=0}^{c} \frac{\binom{D}{d}\cdot\binom{N-n}{n-d}}{\binom{N}{n}} dk_{bw} + (1 - L(p))Dk_{bw} \tag{3}$$

Loss of external lack is found only in situation, when the batch is accepted. This cost embraces the guarantee repairs and loss as a results of forfeiture customers and reputation of company. To assume, that single cost resultant on external loss is equal k_{bz}, this cost, as Kończak [5] shows, is equal:

$$K_{bz}(p) = k_{bz}\sum_{k=0}^{c}(D-k)\cdot P(X=k) \tag{4}$$

Presented structure of the acceptance control cost is correct under the assumption of faultless estimation of quality of the products. Unfortunately, in practice there are no statistical quality control methods, which are error – free. Lavin. M. [6] paid attention on this issue as first one. Since then errors in inspection have been the subject of considerable interest. This problem is dealt with, among others, Kotz and Johnson [4], Sheu et al [8], Chun [1] Wanger [9] and Wanger et al [10].

Considering the fact of occurring classification errors in inspection, an element, which is controlled may be judged as defective, if it is really defective or in case of being correct and being falsely categorized. Similarly, an element, which is controlled may by judged as correct, if it is really correct, or in case of being a defect and being falsely categorized. According to this, the probability of categorized elements as defected is as follows [2]:

$$p^* = p(1-\varepsilon) + q\eta \tag{5}$$

where:

p – a real defectiveness of elements,

ε – a probability of commitment qualification mistake relied on categorize defective element as correct element,

η – a probability of commitment qualification mistake relied on categorize correct element as defective element.

In [2] we can see, that the probability of categorized element as correct is the following:

$$q^* = 1 - p^* \tag{6}$$

In case of occurrence of qualifications errors previously presented expected cost of control, loss of internal and external lacks take the following form presented in [3]:

$$K^*_k(p) = \left[\left(1 - L^*(p)\right)\cdot\left(N - n\right) + n\right]\cdot k_k \tag{7}$$

$$K^*_{bw}(p) = k_{bw}\cdot\left[\left(1 - L^*(p)\right)\cdot D + \sum_{k=0}^{c} k\cdot P\left(X^* = k\right)\right] \tag{8}$$

$$K^*_{bz}(p) = k_{bz}\sum_{k=0}^{c}\left(D - k\right)\cdot P\left(X^* = k\right) \tag{9}$$

where:

$L^*(p)$ – probability of accepting a lot of products in the case of qualifications errors,

X^* – random variable, taking values equal to the number of elements in a sample of the alleged defective.

The acceptance control costs in case of appearance of qualification mistakes should supplement also with cost of redundant repair of elements, which are in fact correct and which were categorized as defects $\left(k_{zn}\right)$. This cost embraces the exchange or repair of correct elements and cost of renewed quality control of those elements after conducted repair. Unit cost of redundant repairs is less than or equal to unit cost associated with external lack in classical reasoning.

4 Alternative classification in the case of many variables

The problem of alternative classification for many variables dealt with Wisniewski K. [11]. He considered a finite sequence of *n* random variables $X_1, X_2, \ldots, X_n$ for which are given probability:

$$\begin{aligned} &P\left(X_i = 1\right) = p_i \\ &P\left(X_i = 0\right) = 1 - p_i \end{aligned} \quad \textit{where} \quad i = 1,2,\ldots,n\,. \tag{10}$$

It is assumed that there are given covariance between pairs of variables $X_1, X_2, \ldots, X_n$ and the central moments of three, four etc. variables $X_1, X_2, \ldots, X_n$.

There are also given a sequence of *n* suitable random variables tacking values 0 and 1 $Y_1, Y_2, \ldots, Y_n$ and two sequences qualifying probabilities of errors of the first and the second type because of the different variables $X_1, X_2, \ldots, X_n$:

$$\begin{aligned} &\alpha_1, \alpha_2, \ldots, \alpha_n, \\ &\beta_1, \beta_2, \ldots, \beta_n. \end{aligned} \tag{11}$$

For random variables $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are led out distributions of sums and products of random variables. There is introduced a random variable $U = \sum_{i=1}^{n} X_i$, that takes values $k = 0,1,2,\ldots,n$ with probabilities $P\left(U = k\right)$.

Similarly there is led out a random variable *V* being qualification variable for the random variable *U*, $V = \sum_{i=1}^{n} Y_i$ taking values $l = 0,1,2,\ldots,n$ with probabilities $P\left(V = l\right)$.

Assuming that the variables of sequence $X_1, X_2, \ldots, X_n$ are the same and their realizations are classified using the same method, there can be regarded as realizations of a random variable *X* with distribution

$P(X=1)=p$. With analogous assumptions regarding the variables of sequence $Y_1, Y_2, \ldots, Y_n$ there obtain the random variable *Y* with distribution. $P(Y=1)=p(1-\alpha)+(1-p)\beta=p'$, where p' is the probability of qualify a random variable *X* as taking a value of *1*.

Using the distribution of products of zero – one random variables, there led out a random variable $Z=\prod_{i=1}^{n} X_i$, which takes the values 1 and 0 with probabilities:

$$\begin{aligned} P(Z=1) &= P(X_1=1, X_2=1, \ldots, X_n=1), \\ P(Z=0) &= 1-P(X_1=1, X_2=1, \ldots, X_n=1) \end{aligned} \tag{12}$$

and a zero – one variable *T* which is a qualifying random variable of variable *Z*. Defining variable *T* there are used the following conditional probabilities:

$$\begin{aligned} P(T=0 \mid Z=1) &= \alpha_{\bar{n}|}, \\ P(T=1 \mid Z=0) &= \beta_{\bar{n}|}. \end{aligned} \tag{13}$$

They represent respectively the probability of error of the first type because of *n* variables, and the probability of error of the second type because of *n* variables.

The probability of error of a first type because of *n* variable $\alpha_{\bar{n}|}$ is associated with the probabilities $\alpha_1, \alpha_2, \ldots, \alpha_n$ of error of a first type because of subsequent variables $X_1, X_2, \ldots, X_n$ by following relation:

$$\alpha_{\bar{n}|} = 1-\prod_{i=1}^{n}(1-\alpha_i). \tag{14}$$

The probability of error of a second type because of *n* variable $\beta_{\bar{n}|}$ is associated with the probabilities $\beta_1, \beta_2, \ldots, \beta_n$ of error of a second type because of subsequent variables $X_1, X_2, \ldots, X_n$ by following relation:

$$\beta_{\bar{n}|} = \prod_{i=1}^{n}(1-\alpha_i+\beta_i)-\prod_{i=1}^{n}(1-\alpha_i). \tag{15}$$

Between the probabilities of error of first and second type because of different variables $X_1, X_2, \ldots, X_n$ there is a following relationship:

$$1-\alpha_{\bar{n}|}+\beta_{\bar{n}|} = \prod_{i=1}^{n}(1-\alpha_i+\beta_i). \tag{16}$$

5 Acceptance inspection for attributes in the case of many controlled features

Presented in the previous section discussion can be transposed to the quality control area. One product may be assessed due to the *n* characteristics. Denoting as "1" a correct element due to the *i* – th feature and marking as "0" a defective element due to this feature, a sequence of *n* zero – one random variables $X_1, X_2, \ldots, X_n$ represents a sequence of information about the quality of the product due to the individual, highlighted features.

In the quality control area, assuming that the probability of not meeting the quality requirements due to for each characteristic in the controlled product is the same (i.e. $\forall_{i\in\{1,2,\ldots,n\}} p_i = p$), the random variable *U* informs due to how many of the *n* characteristics the controlled product is correct, and the random variable *Z* carries information about whether the product is non – defective due to all of the features (*Z = 1*), or if it is defective due to any of the highlighted features (*Z = 0*).

This has particular application in the case of acceptance control carried out because of the more than one feature. Assuming that the batch of products passed to control counts *N* elements, there are *N* – times a sequence of zero – one random variables $X_1, X_2, \ldots, X_n$, and there are *N* – times random variables *U* and *Z*.

From the batch, the sample of the number *m* is selected in a randomly way. The number of defects per one product, denoted by *k*, has a binomial distribution: $P(U=k)=\binom{n}{k}p^k(1-p)^{n-k}$,

where:

n – the number of features, due to which the product is assessed,

p – the defectiveness of the product due to one feature.

The expected number of defects per one product is equal so $w = np$, the expected number of defects per the lot presented to control is equal $w_N = wN = npN$, and the expected number of defects per taken sample consist of *m* products is equal $w_m = wm = npm$

The probability that the product is correct because of all highlighted features is equal $P(Z=1) = P(X_1 = 1, X_2 = 1, \ldots, X_n = 1) = (1-p)^n = q_n$, and the probability that it is defective due to any feature is equal $P(Z=0) = 1 - P(X_1 = 1, X_2 = 1, \ldots, X_n = 1) = 1 - (1-p)^n = p_n$. The probability that a random variable *Z* has took a value of 0 can identify with defectiveness of whole batch passed to inspection.

In the case of incomplete diagnostic skills, the qualification errors committed may be due to any controlled feature. Assuming that the probability of errors due to individual characteristics are the same, the probability of commitment qualification mistake relied on categorize defective element as correct element because of one feature is equal $\underset{i\in\{1,2,\ldots,n\}}{\forall} \varepsilon_i = \varepsilon$, and the probability of commitment qualification mistake relied on categorize correct element as defective element because of one feature is equal $\underset{i\in\{1,2,\ldots,n\}}{\forall} \eta_i = \eta$.

According assumptions being above, the probability of assessment the controlled element as the correct one because of all the highlighted features is equal $P(T=1) = P(X'_1 = 1, X'_2 = 1, \ldots, X'_n = 1) = (-p(1-\eta-\varepsilon)+1-\eta)^n = q'_n$ and the probability that it is defective due to any feature is equal $P(T=0) = 1 - P(X'_1 = 1, X'_2 = 1, \ldots, X'_n = 1) = 1 - (-p(1-\eta-\varepsilon)+1-\eta)^n = p'_n$. The probability that a random variable *T* took the value 0 can identify with defectiveness of whole batch passed to inspection at the qualification errors.

The expected number of defects detected in one element, with incomplete diagnostic skills is equal $w^* = (p(1-\eta-\varepsilon)+\eta)n$.

The idea of acceptance inspection control in the case of assessment the product due to the many features is the same as in the case of assessment it due to one feature, i.e. from batch of products, which counts *N* elements, the sample which counts *m* elements is taken. Any item selected into the sample is subjected to a acceptance inspection, here because of all the *n* distinguished characteristics. Element, in which even one feature does not fit in the designated standards shall be considered as defective. Next, the amount of defective items found in the sample is compared with the acceptance number of used acceptance sampling plan – if defective elements are more than allowed adopted acceptance sampling plan the batch should be considered as defective, otherwise it shall be considered as compatible with the accepted quality requirements.

In the case of assessment items due to more than one feature, it should be remembered that the defectiveness of batch is equal properly p_n in the case of a full diagnostic skills and p'_n in the case of qualifications errors.

5.1 The cost of quality control in case of many controlled features

Presented in the chapter 3 the formulas for the costs of acceptance inspection relate to the classical approach to the acceptance control, i.e. to the case that the product is assessed only due to one feature. In a situation where the batch of products passed to the acceptance control is assessed due to more than one feature, the costs of examination and estimation, loss because of internal and external lack can be expressed by the following formulas:

$$K_k^n(p) = L^n(p)mk_k + (1 - L^n(p))Nk_k \tag{17}$$

$$K_{bw}^n(p) = L^n(p)mnpk_{bw} + (1 - L^n(p))Nnpk_{bw} \tag{18}$$

$$K_{bz}^n(p) = L^n(p)\cdot(N-m)npk_{bz} \tag{19}$$

where:

$L^n(p)$ – the probability of accepting a lot of products in the case of assessed the products due to more than one feature,

m – the sample size,

p – a real defectiveness of single controlled feature,

n – the amount of controlled features,

k_k – the single costs of control of element,

k_{bw} – the single cost of exchange (or repair),

k_{bz} – the single cost of external loss.

The formulas 17 – 19 are correct, assuming that the cost of exchanging or repairing the element due to each feature are the same i.e. the cost associated with internal losses is equal $\underset{i\in\{1,2,\ldots,n\}}{\forall} K_{bw_i} = K_{bw}$, and the cost associated with external losses is equal $\underset{i\in\{1,2,\ldots,n\}}{\forall} K_{bz_i} = K_{bz}$.

With the additional assumption that the diagnostic skills are incomplete, the above formulas take the form of:

$$K_k^{*n}(p) = L^{*n}(p)mk_k + \left(1 - L^{*n}(p)\right)Nk_k \tag{20}$$

$$K_{bw}^{*n}(p) = L^{*n}(p)mnp^*k_{bw} + \left(1 - L^{*n}(p)\right)Nnp^*k_{bw} \tag{21}$$

$$K_{bz}^{*n}(p) = L^{*n}(p)\cdot(N-m)np^*k_{bz} \tag{22}$$

where:

$L^{*n}(p)$ – the probability of accepting a lot of products in the case of assessed the products due to more than one feature and in case of incomplete diagnostic skills,

p^* – the probability of categorized elements as defected.

6 Conclusions

Acceptance control, being the part of the statistical quality control allows to provide the desired level of quality of manufactured products. It can be carried out either when passed to inspection lots of products are assessed only due to one feature, as well as where they are assessed due to more than one feature of the product.

This article moves the topic of multiple alternative qualifications of batches of products presented for inspection. In particular, the formation of the expected costs of acceptance control in case of alternative qualifications of products due to *n* distinguished in the product characteristics, are considered. Considerations carried out assuming the fully diagnostic skills, as well as in case of qualification errors.

Presented patterns on the cost of acceptance inspection conducted due to *n* characteristics of the product, both in case of the complete and the incomplete diagnostic skills, led out on the assumption that defectiveness of the product due to the different characteristics are the same. In the case of incomplete diagnostic skills furthermore assumed that the probability of committed the qualifications errors due to the individual features are the same. In further examination it is planned to extend presented in the work patterns for the case of occurring differences in defectiveness of individual controlled characteristics of product and the different probabilities of committed qualification errors.

Acknowledgements

Supported by the grant No. DEC-2011/03/B/HS4/05630 of the Polish National Science Centre.

References

[1] Chun Y. H.: Economic optimization of off-line inspection procedures with inspection errors. *Journal of the Operational Research Society* ***59*** *(2008)*, 863–865.

[2] Iwasiewicz A.: *Zarządzanie jakością*, PWN, Warszawa – Kraków, 1999.

[3] Iwasiewicz A.: *Zarządzanie jakością w przykładach i zadaniach*, Śląskie Wydawnictwo Naukowe Wyższej Szkoły Zarządzania i Nauk Społecznych w Tychach, Tychy 2005.

[4] Kotz, S., and Johnson, N. L.: Effects of False and Incomplete Identification of Defective Items on the Reliability of Acceptance Sampling. *Mimeo Series No.* ***1386*** (1982).

[5] Kończak G.: *Metody statystyczne w sterowaniu jakością produkcji*. Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice, 2007.

[6] Lavin M.: Inspection Efficiency and Sampling Inspection Plans. *Journal of the American Statistical Association,* ***Vol. 41, No. 236*** *(Dec., 1946)*, 432-438.

[7] Montgomery D. C.: *Introduction to Statistical Quality Control*. John Wiley & Sons Inc., Third Edition, New York, 1997.

[8] Sheu S.H., Chen Y.C., Wang W.Y., Shin N.H.: Economic optimization of off-line inspection with inspection errors. *Journal of the Operational Research Society* ***54*** *(2003)*, 888–895.

[9] Wagn C. H.: Economic off-line quality control strategy with two types of inspection errors. *European Journal of Operational Research* ***179*** *(2007)*, 132–147.

[10] Wang W. Y., Sheu S. H., Chen Y. C., Horng D. J.: On a more general formulation of off-line inspection with inspection errors. *Journal of the Operational Research Society* ***59*** *(2008)*, 865-867.

[11] Wiśniewski K.: O błędach kwalifikacji alternatywnej w przypadku wielu zmiennych. *Przegląd statytyczny zeszyt* ***2*** *(1962)*, 197-217.

Decision Making Support of Logistics Tasks in the Manufacturing System

Bronislav Chramcov[1], Robert Bucki[2]

Abstract. The paper highlights the problem of the manufacturing system in which each machine is equipped with one dedicated tool not being subject to regeneration. Machines are situated within the reach of the robotic arm whose goal is to carry out transport operations of semi-products to the subsequent manufacturing machines in accordance with the determined timescale. Heuristic algorithms are implemented to control the order making process. Minimizing the order making time remains the basic criterion. It is assumed that making orders is carried out either along the determined manufacturing route in case of defining a sequence of manufacturing operations or, in case of not defining such a sequence, according to the determined sequence of operations by means of a heuristic algorithm. The discussed logistic formation is treated as a serial manufacturing system in case of the determined sequence of manufacturing operations. The problem consists in time scaling manufacturing operations for the transporting agent i.e. the robotic arm.

Keywords: manufacturing system optimization, heuristics algorithms, modelling, simulation.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

There is a competitive approach to the problem of managing manufacturing companies as they concentrate on continuous improvements. In the manufacturing environment this means that factory layouts will often be reviewed to make control and management matters more effective. Production organization requires the best possible solutions. Planning must be supported by proper specification where assumptions are to be simplified in order to enable planners to model the exact manufacturing tasks. Then the model forms the basis for building the simulator imitating flow of products in the course of the manufacturing process. The variety and complexity of production processes can be modeled more appropriately by supplying an integrated view of different modeling approaches [2].

Mathematical modeling of the highly complex manufacturing system imitating real production systems must be followed by creating the project based on exact specification assumptions. Then it is the subject to creating software and its subsequent testing. Verification and search for improvements is characteristic for every of these stages. However, for simplicity needs modeling can be carried out in the proposed synthetic manufacturing environment. Extended specification details lead to creating the proper functional model of the analyzed manufacturing system. The most characteristic thing about the problem of replacing a real system by its simplified model is the search for lean solutions minimizing allowable factors observed in manufacturing systems. Modeling the flexibility allows us to adjust the modeled system to the required configuration of production stands resulting from customers' demand e.g. the number of stands in the manufacturing line, implemented tools in each production stand and a sequence of passing ordered elements to be made in other available manufacturing plants of identical production possibilities [3].

The paper focuses on modeling the manufacturing system where orders are given in the order vector. Each element of this vector presents a customer and their specific demand. Each order can be made only from a dedicated charge. The charge is either passed through a determined route in the manufacturing system equipped with machines performing specific operations on the semi-product or it is made by machines which are determined with the use of the heuristic algorithm. It is also assumed that there is the buffer store zone serving storing needs. The capacity is defined and there is an option of switching it off in case of the lack of storing space. Moreover, specification and the subsequent software assume the possibility of multiplying the discussed manufacturing system in a parallel way according to the needs. To satisfy the lean approach needs, heuristic algorithms choose

[1] Tomas Bata University in Zlín, Faculty of Applied Informatics, Nad Stráněmi 4511, 760 05 Zlín, The Czech Republic, chramcov@fai.utb.cz.
[2] Institute of Management and Information Technology, ul. Legionów 81, 43-300 Bielsko-Biała, Poland, rbucki@wsi.net.pl

the product to be made to meet the stated criterion. The operating principle forms the basis for creating the simulator of the modeled manufacturing system. However, there are production strategies which decide about the moment of beginning making elements of the order vector.

Manufacturing systems require optimization where a wide range of methods and approaches are implemented. The basic assumption requires creating adequate models [4] with the use of e.g. the process oriented approach [9], the value oriented approach [10], the multi-agent approach [11], the discrete or continuous Petri-nets approach [7], the fuzzy-multi objective approach [5] or heuristic algorithms which are responsible for meeting the set criterion [1], [8]. Optimization of manufacturing logistics as a complex of all direct and indirect activities plays a more and more important role in its relation with manufactured goods, charges, implemented technologies and production organization [6].

2 Specification of the manufacturing system

It is assumed that the manufacturing system consists of the charge zone, manufacturing hall, storing place for tools, storing place for semi-products and dispatching zone from which products are passed to customers. In the assumed real system machines are placed at different locations in the manufacturing hall in accordance with the manufacturing arrangement. Each machine performs only one type of manufacturing operation. Charges are assumed to be supplied on the demand basis with no delay. Each type of product is associated with a certain type of charge.

Let us assume that customers set orders in accordance with the vector of orders in the form (1) where z_n^k is the number of units of the *n*-th production order at the *k*-th stage. The stage *k*, *k*=1,…,*K* is the moment of making the production decision.

$$Z^k = \left[z_n^k\right] \tag{1}$$

It is further assumed that there is a robotic arm which is responsible for the following operations:

- taking charge material from the entrance gate to the determined machine,
- transporting semi-products from a preceding machine to the determined subsequent one,
- transporting the ready product to the dispatching gate,
- transporting semi-products to the storing zone and back,
- manipulating with semi-products in the storing zone,
- replacing a used tool with a new one.

It is assumed that the manufacturing system is equipped with a buffer zone where semi-products are stored for the determined period of time if required in case of the lack of manufacturing capacities for the certain semi-product. The buffer zone is located within the reach of the robotic arm.

2.1 Mathematical description of the manufacturing system

Let us assume that, from the information point of view, the manufacturing system is brought to its simplified form in which machines are to be associated with a certain defined point. This leads to creating the structure consisting of *I* rows and *J* columns. From this point of view it is possible to define the structure matrix of the system in the form (2) where the element $e_{i,j}$ takes the value of one if the machine is placed in the *i*-th row of the *j*-th column, otherwise the element $e_{i,j}$ takes the value of zero.

$$E = \left[e_{i,j}\right] \tag{2}$$

Let us assume that there are *A* types of tools which can perform *B* manufacturing operations which is shown in the matrix form (3) where $\theta_{a,b}$ represents the *a*-th tool able to perform the *b*-th operation. At the same time $\theta_{a,b} = 1$ if the *a*-th tool can perform the *b*-th operation, $\theta_{a,b} = 0$ otherwise.

$$\Theta = \left[\theta_{a,b}\right] \tag{3}$$

When a tool needs to be replaced, the machine is excluded from the manufacturing process. It is assumed that the robotic arm is equipped with the required function enabling it to carry out the adequate replacement operation. It is assumed that a certain machine is equipped with a dedicated tool only which is shown by means of the adjustment matrix in the form (4). The variable $h_{i,j}$ specifies the number of tool (1 to *A*) in the machine

placed in the *i*-th row of the *j*-th column. If the machine is not placed in the *i*-th row of the *j*-th column, then $h_{i,j} = 0$.

$$H = \left[h_{i,j}\right] \tag{4}$$

The base life matrix of tools used in the manufacturing system for a new brand set of tools used to manufacture units of the order is given in the form (5) where $g_{n,a}$ is the base number of units of the *n*-th order which can be manufactured before the *a*-th dedicated tool is completely worn out and requires an immediate replacement.

$$G = \left[g_{n,a}\right] \tag{5}$$

Additionally, it is possible to specify a manufacturing operation which can be carried out in the given machine. Therefore, the matrix of operations is introduced in the form $O = \left[O_{i,j}\right]$ where $o_{i,j}$ specifies the number of a manufacturing operation which can be carried out in the machine placed in the *i*-th row of the *j*-th column.

Manufacturing operations on order units are carried out in various numbers of machines in different sequences. Let us introduce the vector of route (for the case when the order of manufacturing operations is obligatory) in the form (6). This vector specifies the manufacturing sequence for the *n*-th order (the sequence of manufacturing operations). The element $d(n)_u$ represents the number of manufacturing operation (1 to *B*) which is carried out on the *u*-th order. This vector has a variable length. It depends on the number of manufacturing operation which has to be carried out on the specific order unit.

$$D(n) = \left[d(n)_u\right] \tag{6}$$

2.2 Definition of transportation and replacement times

It is possible to introduce matrices of transportation times for the *n*-th type of product. The matrix in the form (7) specifies the transportation times between the machines in the manufacturing system where $\tau^{TR}(n,i,j)_{i',j'}$ is the transportation time (for the *n*-th order) from the machine placed in the *i*-th row of the *j*-th column to the machine placed in the row *i'* of the column *j'*. At the same time: $\tau^{TR}(n,i,j)_{i',j'} > 0$ if there is a transport operation taken into account, $\tau^{TR}(n,i,j)_{i',j'} = 0$ otherwise.

$$T^{TR}(n,i,j) = \left[\tau^{TR}(n,i,j)_{i',j'}\right] \tag{7}$$

Similarly, it is possible to define the matrix of transportation times from the charge zone as well as the matrix of transportation times to the dispatching zone. These matrices take the form (8). The variable $\tau^{IN}(n)_{i,j}$ specifies the transportation time (for the *n*-th order) from the charge zone to the machine placed in the *i*-th row of the *j*-th column and the variable $\tau^{OUT}(n)_{i,j}$ specifies the transportation time (for the *n*-th order) from the machine placed in the *i*-th row of the *j*-th column to the dispatching zone.

$$T^{IN}(n) = \left[\tau^{IN}(n)_{i,j}\right] \qquad T^{OUT}(n) = \left[\tau^{OUT}(n)_{i,j}\right] \tag{8}$$

The matrix of times of replacement operations carried out by the robotic arm is introduced in the form (9) where $\tau^{R}_{i,j}$ is the replacement time of the specified tool (which is defined by the matrix H) in the machine placed by the robotic arm in the *i*-th row of the *j*-th column.

$$T^{R} = \left[\tau^{R}_{i,j}\right] \tag{9}$$

Moreover, there is a need to introduce the matrix of transportation times for new tools which are stored in the storing place for tools. The matrix in the form (10) is defined where $\tau^{Tool-IN}_{i,j}$ is the transportation time of a new tool from the storing place to the machine placed in the *i*-th row of the *j*-th column.

$$T^{Tool-IN} = \left[\tau^{Tool-IN}_{i,j}\right] \tag{10}$$

Additionally, there is a need to introduce the matrix of transportation times for worn out tools to the storing place for used tools. This matrix is presented in the form (11) where $\tau_{i,j}^{Tool-OUT}$ is the transportation time of a worn out tool from the machine placed in the *i*-th row of the *j*-th column to the storing place for worn out tools.

$$T^{Tool-OUT} = \left[\tau_{i,j}^{Tool-OUT}\right] \tag{11}$$

It is assumed that the manufacturing system is equipped with a buffer zone where semi-products of the *n*-th order are stored for the determined period of time if required in case of the lack of manufacturing capacities for the certain semi-product. The buffer zone is located within the reach of the robotic arm. The matrix of transport times of semi-products to the determined place in the buffer zone is introduced in the form (12) where $\tau^{TR-Buf}(n)_{i,j}$ is the transport time of the *n*-th type semi-product to the determined place in the buffer zone from the machine placed in the *i*-th row of the *j*-th column. Additionally, it is assumed that the matrix of transport times of semi-products to the determined machine from its storing place in the buffer zone is also defined in the form (12).

$$T^{TR-Buf}(n) = \left[\tau^{TR-Buf}(n)_{i,j}\right] \tag{12}$$

Finally, the matrix of allocation operation times in the buffer zone is presented in the form $T^{Buf} = \left[\tau_n^{Buf}\right]$ where τ_n^{Buf} is the allocation operation time of the *n*-th semi-product in the buffer zone.

2.3 State of the manufacturing system

Throughout the course of manufacturing process the state of the system changes. The state of the manufacturing system changes after every operation on the order unit. Therefore, there is a need to analyze the state of the order, tools in machines, charge materials, storing zones, etc.

The state of the order changes after each production decision. The order vector is modified at each *k*-th state according to the form (13) where x_n^k is the number of units of the *n*-th order made at the *k*-th state.

$$z_n^k = \begin{cases} z_n^{k-1} - x_n^k & \text{if a certain number of units of the } n-\text{th order is made at the } k-\text{th state} \\ z_n^{k-1} & \text{otherwise} \end{cases} \tag{13}$$

It is possible to present the state of the tool in machines by means of the matrix of state of the manufacturing system in the form $S^k = \left[s_{i,j}^k\right]$ where $s_{i,j}^k$ represents the value of relative tool wear ($0 \le s_{i,j}^k \le 1$) of the tool in the machine placed in the *i*-th row of the *j*-th column. This element takes its value according to (14) where $x^k(n)_{i,j}$ is the number of units of the order z_n made by the machine placed in the *i*-th row of the *j*-th column and where $g_{n,a}$ is the base number of units of the *n*-th order which can be manufactured before the *a*-th tool (the tool in the specified machine) is completely worn out and requires an immediate replacement.

$$s^k = \begin{cases} s_{i,j}^{k-1} + \dfrac{x^k(n)_{i,j}}{g_{n,a}} & \text{if a certain number of units of the } n-\text{th order is made} \\ & \text{in the } i-\text{th machine in the } j-\text{th column at the } k-\text{th state} \\ s_{i,j}^{k-1} & \text{otherwise} \end{cases} \tag{14}$$

If the state of the tool should be exceeded ($s_{i,j}^k \ge 1$), no unit of any order can be made in the specified machine and it triggers the need to carry out the replacement process of the tool. If the tool in the machine placed in the *i*-th row of the *j*-th column has to be replaced with a new one, the state of this tool takes the value zero after carrying out the replacement procedure.

In addition, it is useful to monitor how many conventional units of any order can be made with the use of the specific machine. Hence, the matrix of the available capacity of the manufacturing system in the form $P^k(n)_{i,j} = \left[p^k(n)_{i,j}\right]$ is introduced where $p^k(n)_{i,j}$ is the number of conventional units of the *n*-th order which still can be realized in the *i*-th machine of the *j*-th column.

If there is a remaining available capacity in the machine in the *i*-th row of the *j-t*h column but the subsequent unit of the *n*-th order cannot be made fully in this machine, then the replacement process is carried out automatically. The available capacity of the machine in the *i*-th row of the *j*-th column for the *n*-th order is determined in the form (15) where expression $\lfloor x \rfloor$ represents the floor function and $g_{n,a}$ is the base number of units of the *n*-th order which can be manufactured before the *a*-th tool (the tool in the specified machine) is completely worn out and requires an immediate replacement.

$$p^k(n)_{i,j} = \left\lfloor g_{n,a} \cdot \left(1 - s^k_{i,j}\right) \right\rfloor \tag{15}$$

3 Manufacturing approaches and heuristics control algorithms

Each order may be associated with the generated sequence of manufacturing operations where defined tools require adjustment of operation times. In the discussed case, the manufacturing sequence for the *n*-th order is defined by the vector *D(n)*. However, there are three manufacturing approaches to be discussed (from the point of view of operation sequence).

1. The first method assumes that the sequence of manufactured orders (products) remains unchanged throughout the whole manufacturing process i.e. the manufacturing sequence is determined. An element of the order matrix to be made is chosen with the use of heuristic algorithms (discussed further). In this case the system is treated as the serial manufacturing system and there is a need to implement the vectors of routes.
2. The second method is based on the principle that any order can be made without the need to follow any sequence restriction of manufacturing operations (the sequence is optional) i.e. a machine is chosen according to a specified heuristic algorithm in order to meet the defined criteria.
3. The third method mixes the two previous methods as some orders require a defined manufacturing sequence whereas others do not.

In case of the first approach the sequence of operations on the *n*-th order is predefined and the order of making products cannot be alternated. Searching for solutions minimizing the total order making time consists in minimizing the standstill time of the manufacturing system as a whole or its individual elements (machines). Moreover, the manufacturing time cannot be increased so there is a need to search for other solutions such as e.g. minimizing the capacity loss of tools. There is also a need to recalculate the life of tool for the specific type of operation i.e. the one performed on the specific type of order.

Additionally, there are two manufacturing approaches from the point of view of starting manufacturing the next order:

1. The manufacturing process continues according to the chosen heuristic algorithm until all units of order chosen for production are made and only then the next determined order is subject to manufacturing.
2. It is assumed that exactly when a certain *n*-th order frees the manufacturing space another order enters the manufacturing zone. The continuity of the manufacturing process is guaranteed by using every possible machine so a decision about manufacturing is made one after another. The priority is to put charge material into the machine first. The robotic arm can make all other transportation tasks after delivering the charge material to all machines. Only then can semi-products be transported to the subsequent determined machines.

Finally, it is possible to change the sequence of orders on the basis of their priority. In this case the matrix of order priority must be introduced in the form $\Pi = [\pi_n]$ where π_n specifies the priority of the manufacturing task for the *n*-th order. The priority element takes the integer values where $\pi_n = 0$ means the lowest priority.

3.1 Heuristic algorithms

In order to control the choice of the order and subsequently directing it to the manufacturing zone we need to implement heuristics which determine elements from the vector *Z* for the production process. In this case it is possible to put forward one of available heuristic algorithms from a wide range of them. Sample heuristic algorithms (16) are considered for demonstration purposes however, their number may be much bigger. For example, the first heuristic chooses the order characterized by the maximal number of units to be made.

Let z^k_μ be the element to be made in the manufacturing zone at the *k*-th stage, $1 \le \mu \le N$. On this basis the heuristic algorithms (16) are put forward.

$$z^k_\mu = \max_{1<n<N} z^k_n \qquad z^k_\mu = \min_{1<n<N} z^k_n \qquad z^k_\mu = \max_{1<n<N} \frac{z^k_n}{\sum_{n=1}^{N} z^k_n} \qquad z^k_\mu = \min_{1<n<N} \frac{z^k_n}{\sum_{n=1}^{N} z^k_n} \tag{16}$$

In order to control the choice of the specific machine heuristics which determine the machine for making the unit of order on the basis of the available capacity of the manufacturing system are introduced. The algorithms of the maximal and minimal available capacity of the system can be put forward. The algorithm of the maximal available capacity of the system chooses the machine with the tool which can make the biggest number of order units. On the other hand, the algorithm of the minimal available capacity chooses the machine with the most worn out tool.

Some manufacturing criteria can be used to evaluate implemented control algorithms. In our case, the total order realization time, the value of the unused capacity of tools or the total tool replacement time is taken into account.

4 Conclusion

Mathematical modeling of such a complex manufacturing environment results in minimizing the number of chaotic factors which may disturb the proper functioning of the system. However, only real implementation of solutions presented in the paper may prove the correctness of the modeling approach emphasized hereby. Nevertheless, without doubt, minimizing production losses, maximizing manufacturing efficiency and, finally, optimizing production procedures may result in decreasing manufacturing costs i.e. cutting energy costs, labor force costs, etc.

Minimizing the order making time remains the main criterion to be discussed; however, verification of order making is to be carried out from the point of view of minimizing costs which should result in creating a two- or three-criterion model. On the basis of practical observations of manufacturing processes times of operations must be modified in order to make calculations more precise. Creating a simulation tool on the basis of the above assumptions will enable the authors to carry out simulation experiments, analysis of which should result in adequate conclusions.

Acknowledgements

This work was supported in part by the European Regional Development Fund within the project CEBIA-Tech No. CZ.1.05/2.1.00/03.0089.

References

[1] Bucki, R., and Suchánek, P.: The Method of Logistic Optimization in E-commerce. *Journal of Universal Computer Science* **18** (10, 2012), 1238-1258.
[2] Bucki, R., Suchánek, P., and Vymětal, D.: Information Modelling of the Complex System of the Parallel Manufacturing Process. In: *Recent Researches in Automatic Control and Electronics. Proceedings of 14th WSEAS Int. Conf. on Automatic Control, Modelling & Simulation (ACMOS '12)*, 2012, pp. 43-48.
[3] Bucki, R., Suchánek, P., and Vymětal, D.: Information Control of Allocation Tasks in the Synthetic Manufacturing Environment. *International Journal of Mathematics and Computers in Simulation* **6** (3, 2012), 324-332.
[4] Chramcov, B., and Balate, J.: Model-building for Time Series of Heat Demand. In *Annals of DAAAM for 2009 & Proceedings of the 20th International DAAAM symposium*, 2009, pp. 195-196.
[5] Dotoli, M., Fanti, M.P., Mangini, A.M. and Tempone, G.: Fuzzy Multi-objective Optimisation for Network Design of Logistic and Production Systems. In: *ETFA 2005: 10th IEEE International Conference on Emerging Technologies and Factory Automation*, 2005, pp. 475-482.
[6] Farahani, R., Rezapour, S.and Kardar, L.: *Logistics Operations and Management: Concepts and Models* (Elsevier Insights). Elsevier; 1 edition, 2011.
[7] Macias, E.J., and de la Parte, M.P. Simulation and Optimisation of Logistic and Production Systems Using Discrete and Continuous Petri Nets. *Simulation-transactions of the Society for Modeling and Simulation International* **80** (3, 2004), pp. 143-152.
[8] Modrák, V.: Using the Simulation Technique for Planning the Industrial Production in Metallurgy. *Metalurgia International* **18** (2, 2013), pp. 17-20.
[9] Řepa, V.: *Podnikové procesy: Procesní řízení a modelování*. Praha: Grada Publishing, 2006.
[10] Shapiro, J. F.: *Modeling the Supply Chain*. South-Western CollegePub, 2006, 618 p.
[11] Wooldridge, M.: *An Introduction to Multi Agent Systems*. Wiley, 2009, 484 p.

Optimization of Logistics in Service Company for Sustainability Performance

Zuzana Chvatalova[1], Jiri Hrebicek[2], Hana Zaoralova[3]

Abstract. The paper presents the method of operational research for the optimization of logistics in a service company. This is important for the decision making of company management related to logistic system and for the support of its sustainable business performance. There is described the relevant transportation problem, designed and solved its mathematical model. Information and communication technology tools provide an efficient environment for the solving mathematical models. Without this support it would be impractical to solve large logistic tasks due to the computational complexity, extensive sets of variables, etc. The Maple system is used for the solution of transport problem. Its Optimization Package is briefly presented. The solution of transport problem is interpreted as economic outputs, which are understandable and important for company management. The paper links to one of co-author thesis defended on the Faculty of Business and Management of the Brno University of Technology. This research can be the stimulus for the cooperation of universities with companies and create a modern uncomplicated technological support for solving problems and decision making in logistics.

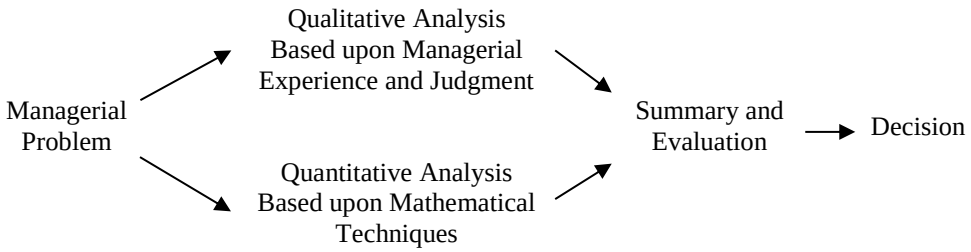
Keywords: operational research, logistics, management, Maple, sustainability.

JEL Classification: C44, C88, D21
AMS Classification: 68N01, 90C05, 91B74

1 Introduction

Current business processes in enterprises tend to be complex because of distance and time constraints, parallel activities, complex structure of the organization and processes, etc. This can lead to inefficiency in the use of production factors and thereby can endanger the corporate sustainability performance (CSP). It is currently closely observed and supported globally.

The management of each organization is responsible for the identification of problems and for finding solutions. Intuition, experience and knowledge help to find solutions. The current information and communication technologies (ICT) help to implement advanced mathematical methods in practice and facilitate the cooperation between businesses and academic institutions. This is a valuable help for decision makers in organizations. Various authors variously describe decision-making processes. For example in [11] we can see Scheme 1.



Scheme 1 Decision-making process (Source: [1])

Approach through an operational research means to gain scientifically optimized solutions for the real implementation in practice, how we can see more in detail for example in [2], [6].

Before verification, application and obtaining of feedback on the results it is necessary to:

[1] Brno University of Technology, Faculty of Business and Management, Kolejni 4, 612 00 Brno, Czech Republic e-mail: chvatalova@fbm.vutbr.cz.
[2] Mendel University in Brno, Faculty of Business and Economics, Zemedelska 1, 61300 Brno, Czech Republic e-mail: hrebicek@mendelu.cz.
[3] Brno University of Technology, Faculty of Business and Management, Kolejni 4, 612 00 Brno, Czech Republic e-mail: e-mail: zaora01@std.fbm.vutbr.cz.

- identify and define the problem in the economic/business environment and operations that must be included in its solving (whole problem has to be well understood), and determine an *economic model*;
- design and describe its *mathematical model* (select relevant variables, appropriate methods and techniques for its solution);
- choose the correct *applied software* for solving this mathematical model and *solve this mathematical model*;
- *evaluate and interpret results* - from mathematical variables to economic indicators understandable to the company management.

In [3] authors say: "*Most models simplify reality. Generally, model can only approximate real-world behavior. One very powerful simplifying relationship is proportionality*."

The advantages of the mathematical model include:

- notation is formalized - historical development lead to the formalization;
- precise rules for calculations, variables and structures are given;
- possibility of using ICT for its solution [4].

2 Linear Programming as a Support of Sustainability Performance in Logistics of Service Company

2.1 Corporate Sustainability Performance

CSP and its monitoring are currently a key factor in the process of sustainability development all over the world. Developed methods of measurement, evaluation, reporting, and subsequent decision-making and corporate management performance are founded on the base of several selected environmental, social, corporate governance (ESG), and economic indicators, more in [8]. Logistics in a service company is very important from possibilities of optimizing and innovating processes. Relevant indicators can be used to measure CSP (they have more favorable values) and help the company to achieve a sustainable business growth.

2.2 Economic Problem, Economic Model

The analyzed service company belongs among the large Czech enterprises according to the classification of the Czech Statistical Office. According to [11] this company (based in 1993) imports and exports fresh foods from around the globe and provides complete comprehensive logistic services. Its activities are based on knowledge management (for example, in biological processes, logistics, transport, human resources, foreign trade, etc.). It covers four cities in the Czech Republic and Bratislava in the Slovakia. The company uses the enterprise information system (IS) "Enterprise Resource Planning (ERP) *Karat"*, which integrates financial, economic, business, and warehousing software [7]. Company strengths include continuous innovations, the use of modern technologies and own transport fleet. The main strategic goal is to build the prosperous competitive company with national coverage wholesale viable within the European Union (EU). Short-term goals include increasing efficiency, productivity, and performance, more in [11].

Therefore the company management is interested in optimization of logistics processes leading to lower costs of operations in its warehouse. These issues belong to the area of operations research, linear programming (LP). This problem has been solved by the principle of task distribution, as a transportation problem [6].

Specifically, the management solves the question: how to load up four trucks K, L, M, N with pallets from three stations 1, 2, 3 in the warehouse. The desired objective is the most efficient loading sequence. The spatial distribution of trucks and stations is shown in Figure 1.

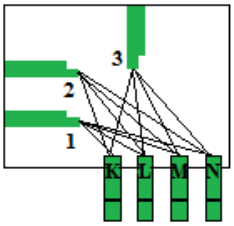


Figure 1 Material flows (Source: [11])

To define a transportation problem, firstly, we have to evaluate and define the objective, particularly a work performance in the warehouse. For this purpose a process was selected so called *the import truck acceptance*, which is described and formalized in the internal enterprise guideline, more in [11]. This process was gradually monitored. The costs of sub-processes (i.e. partial operations of this process) have been calculated. Therefore,

the times and distances were measured (i.e., times and distances, in which is realized the reception of goods and its unloading from arrival of the truck to the storage of last pallets into right place, it includes also the selection of storage - considering an appropriate operating temperature, administration, quality control, etc.). The truck content and the process of unloading were closely monitored. Subsequently, the work with the pallet trucks could be clearly valued.

	Value	Unit		Value	Unit
Worker hourly wage	120	CZK/h	**Pallet truck price**	13.125	CZK
Pallet truck hourly "wage"	22.5	CZK/h	**Total price**	83.125	CZK
Operation time	35	min	**Price per minute**	2.375	CZK/min
Operation distance	3773.4	m	**Price per meter**	0.022029	CZK/m
Worker price	70	CZK	**Price meter per minute**	107.8114	m/min

Table 1 Monitored values (Source: Authors´ elaboration according to [11])

Note: CZK (Czech Crown); m (meter); min (minute); h (hour); pallet truck (accumulator forklift truck); worker hourly wage (it is generally given); pallet truck "wage" (one worker and one pallet truck is needed for the whole process of unloading for one truck); operation time and operation distance (for one worker + one pallet truck needed for unloading of one truck); pallet truck hourly "wage" (this imaginary pallet truck "wage" was derived on the basis of the price for the consumed energy of this truck and its amortization); operation time/distance (observed manipulation time/distance of worker + pallet truck); worker/pallet truck price (related wage to operation time); total price (worker price + pallet truck price); price per minute/per meter (total price related to operation time/operation distance); meter per minute (ratio = price per meter related to price per minute).

Subsequently, distance matrix D and value (price) matrix V were introduced, where values $d_{i,j}$ and $v_{i,j}$ (i = 1, 2, 3, j = K, L, M, N) are distances and prices related to i-station and j-truck, unit for distance is meter and unit for prices is CZK:

$$D := \begin{bmatrix} d_{1,K} & d_{1,L} & d_{1,M} & d_{1,N} \\ d_{2,K} & d_{2,L} & d_{2,M} & d_{2,N} \\ d_{3,K} & d_{3,L} & d_{3,M} & d_{3,N} \end{bmatrix} = \begin{bmatrix} 16 & 21 & 26.5 & 32 \\ 27 & 29.5 & 34 & 38.5 \\ 23.5 & 25.5 & 29 & 32.5 \end{bmatrix} \tag{1}$$

$$V := \begin{bmatrix} v_{1,K} & v_{1,L} & v_{1,M} & v_{1,N} \\ v_{2,K} & v_{2,L} & v_{2,M} & v_{2,N} \\ v_{3,K} & v_{3,L} & v_{3,M} & v_{3,N} \end{bmatrix} = 0.022029 \cdot \begin{bmatrix} 16 & 21 & 26.5 & 32 \\ 27 & 29.5 & 34 & 38.5 \\ 23.5 & 25.5 & 29 & 32.5 \end{bmatrix} = \begin{bmatrix} 0.3524640 & 0.4626090 & 0.5837685 & 0.7049280 \\ 0.5947830 & 0.6498555 & 0.7489860 & 0.8481165 \\ 0.5176815 & 0.5617395 & 0.6388410 & 0.7159425 \end{bmatrix} \tag{2}$$

The trucks capacity T and the stations capacity S are given (number of pallets) in (3), (4):

$$T := \begin{bmatrix} capK & capL & capM & capN \end{bmatrix} = \begin{bmatrix} 33 & 36 & 30 & 33 \end{bmatrix} \tag{3}$$

$$S := \begin{bmatrix} cap1 & cap2 & cap3 \end{bmatrix} = \begin{bmatrix} 42 & 51 & 39 \end{bmatrix} \tag{4}$$

The summary of inputs of transportation problem is given in Tab. 2.

	K	***L***	***M***	***N***	**Capacity**
1	0.3524640	0.4626090	0.5837685	0.7049280	42
2	0.5947830	0.6498555	0.7489860	0.8481165	51
3	0.5176815	0.5617395	0.6388410	0.7159425	39
Requirements	33	36	30	33	132

Table 2 Input table for the transportation problem formulation (Source: Authors´ elaboration according to [11])

2.3 Mathematical Model of Transportation Problem

Let us consider the transportation problem and describe it in (5), (6) and (7) with input values in Tab. 2 with 12 variables x_{ij} (i=1, 2, 3; j=1, 2, 3, 4) (the individual numbers of pallets that are transported between the station and trucks on the individual routes), where we substitute letters K, L, M, N describing trucks by integers 1, 2, 3, 4. These integers we will use in Maple computations. The solution is required in integer form. Below is its *standard form* of the transportation problem in the LP formulation [6].

Minimized objective function:

$$\begin{aligned} &0.3524640x_{11} + 0.4626090x_{12} + 0.5837685x_{13} + 0.7049280x_{14} + \\ &+ 0.5947830x_{21} + 0.6498555x_{22} + 0.7489860x_{23} + 0.8481165x_{24} + \\ &+ 0.5176815x_{31} + 0.5617395x_{32} + 0.6388410x_{33} + 0.7159425x_{34} \end{aligned} \tag{5}$$

with seven linear equality constraints:

$$\begin{aligned} x_{11} + x_{12} + x_{13} + x_{14} &= 42 \\ x_{21} + x_{22} + x_{23} + x_{24} &= 51 \\ x_{31} + x_{32} + x_{33} + x_{34} &= 39 \\ x_{11} + x_{21} + x_{31} &= 33 \\ x_{12} + x_{22} + x_{32} &= 36 \\ x_{13} + x_{23} + x_{33} &= 30 \\ x_{14} + x_{24} + x_{34} &= 33 \end{aligned} \tag{6}$$

and non-negativity bounds:

$$x_{ij} \geq 0, \ (i=1, 2, 3; \ j=1, 2, 3, 4) \tag{7}$$

2.4 Optimization Package of Maple

The Maple system was chosen for solving the problem (5), (6) and (7). It includes the *Optimization Package* for solving optimization tasks. This package has commands which allow the user to find the optimum of various types of functions of many variables with various types of constraints. A lot of information, instructions, options, examples, calling sequence, notes, etc. is clearly described on the Maple Help Pages or Maple Online Help,[9] [10]. Even if it is very powerful and complex, it is still easy to use Maple for new users.

The Optimization package consists of subpackages: *Optimization* [*ImportMPS*] for import of math program data file; *Optimization* [*Interactive*] for display of interactive interface in the Optimization package; *Optimization* [*LPSolve*] for solving linear programming problems; *Optimization* [*LSSolve*] for solving least-squares problems; *Optimization* [*Minimize*] for finding minimum of an objective function, possibly subject to constraints; *Optimization* [*Maximize*] for finding maximum of an objective function, possibly subject to constraints; *Optimization* [*NLPSolve*] for solving of nonlinear programming problems; *Optimization* [*QPSolve*] for solving quadratic programing problems. More advanced Matrix form of input is described for *LPSolveMatrixForm*, *LSSolveMatrixForm*, *NLPSolveMatrixForm*, *QPSolveMatrixForm*. Matrix form leads to more efficient computation, but is more complex, more in [9], [10].

2.5 Solving Transportation Problem in Maple

We solved the mathematical model (5), (6) and (7) of above transport problem with Maple commands from the *Optimization*[*LPSolve*] package and using the manual [5].

The Maple Online Help [9] describes: "*The LPSolve command solves a linear program (LP), which involves computing the minimum (or maximum) of a linear objective function subject to linear constraints. Continuous, integer, mixed-integer and binary (or zero-one) LPs can be solved. Throughout this page, the term integer programming is used to represent all forms of integer programs. Integer programs are specified using the assume, integer variables or binary variables options. Maple returns the solution as a list containing the final minimum (or maximum) value and a point (the extreme). If the output = solutionmodule option is provided, then a module is returned. Integer programs are solved using a branch-and-bound strategy, in which a tree of nodes is created. Each node corresponds to a continuous LP subproblem, which is solved using the active-set method*."

In Figure 2 the Maple worksheet lists all computations.

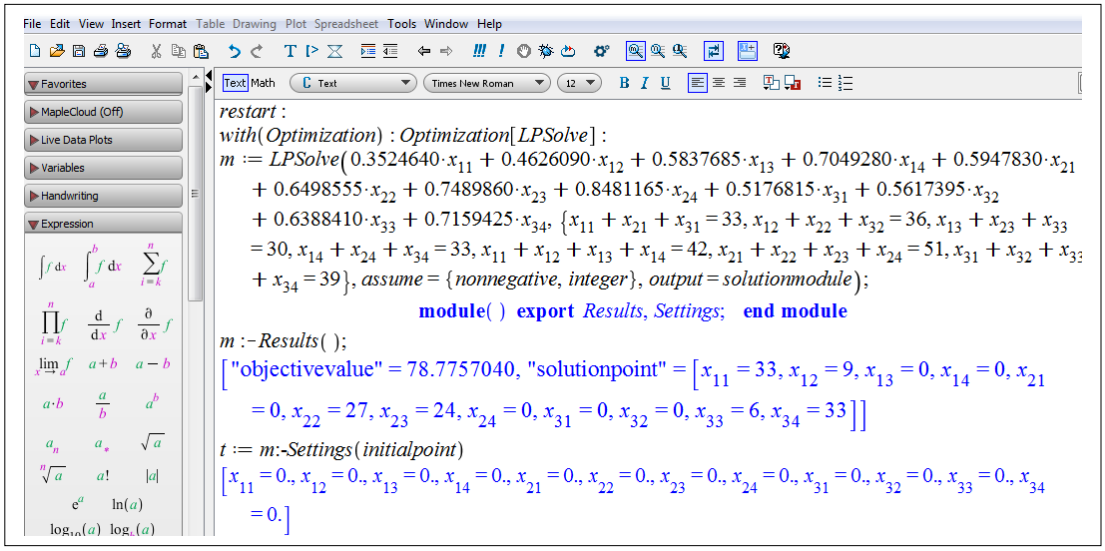


Figure 2 Maple worksheet of solved mathematical model (5), (6) and (7)
(Source: Authors´ elaboration according to [11])

We can see from the Maple computations at worksheet (Fig. 2) that the minimum of LP task is equal to 78.7757040 CZK.

Thus, 33 pallets must be moved into the truck *K* and 9 pallets into the truck *L* from the station 1; 27 pallets must be moved into the truck *L* and 24 pallets into the truck *M* from the station 2; 6 pallets must be moved into the truck *M* and 33 pallets into the truck *N* from the station 3 at process costs approximately 78.78 CZK respecting the input conditions. Exactly 3 + 4 – 1= 6 values $x_{i,j}$ are positive (non-degenerate solution condition).

Note, if we don´t use in the command *LPSolve* the integer assumption, the Maple system returns a very similar solution values, which rounded is identical to values at Fig. 2.

3 Conclusion

Currently interaction between theory and practice plays a key role. The Faculty of Business and Management of the Brno University of Technology (FBM BUT) strongly supports this trend. Academics and students have the conditions for using modern technologies and possibilities of access to scientific resources beyond teaching and research. This paper links to one of co-author thesis realized on the Faculty of BUT, which is based on this trend. Thesis [11], which was the inspiration for this paper, was based on this trend. The paper presented the method of operational research for the optimization of logistics in the service company. This is important for the decision making of company management related to logistics system and for the support of the company sustainable business growth. There is described the relevant transportation problem, designed and solved its mathematical model. The mathematical modelling helps to solve specially described economic problems in FBM BUT courses. Results are interpreted as economic outputs, which are understandable and important for company management. The Maple system is also used for this research. The new version of the Maple system simplifies the solving of various problems in the real world. The presented research approach with using applied software helps to increase the corporate sustainability performance to ensure its sustainable development and helps to optimize its business activities in logistic process. This research stimulates the cooperation of universities with companies and creates a modern uncomplicated technological support for solving problems and decision making in logistics.

Acknowledgements

This paper is supported by The Czech Science Foundation. Name of the Project: Measuring Corporate Sustainability in Selected Sectors. Reg. No. 14-23079S.

References

[1] Anderson, D. R., et al.: *An Introduction to Management Science: Quantitative Approaches to Decision Making*. 5th ed. West Publishing Company, USA, 1991.
[2] Fiala, P.: *Modely a metody rozhodování*. Oeconomica, Vysoká škola ekonomická v Praze, Praha, 2013.

[3] Giordano, F. R, Fox, W.P., Horton, S.B. and Weir, M. D.: *A First Course in a Mathematical modeling*. 4th ed., international student edition. Brooks/Cole Cengage Learning, Canada, 2009.
[4] Hřebíček, J., Pospíšil Z. and Urbánek J.: *Úvod do matematického modelování s využitím Maple.* [online]. Institut biostatistiky a analýz MU, Brno, 2010. http://www.iba.muni.cz/esf/res/file/Uvod_do_MM_el_verze.pdf (last accessed 15.04.14.).
[5] Chvátalová, Z.: *Malý Maple manuál*. [online]. Fakulta podnikatelská Vysokého učení technické v Brně, Brno, 2011. http://www.maplesoft.cz/sites/default/files/img/manual_ chvatalova.pdf (last accessed 15.04.14.).
[6] Jablonský, J.: *Operační výzkum*. 2. vyd. Vysoká škola ekonomická v Praze, Praha, 1998.
[7] *Karat - Informační systém.* [online]. Komplexní ERP podnikový informační systém KARAT. KARAT Software a. s. 2006 – 2014. http://www.karatsoftware.cz/ (last accessed 15.04.14.).
[8] Kocmanová, A., Hřebíček, J., Dočekalová, M., Hodinka, M., Hornungová, J., Chvátalová, Z., Kubálek, T., Popelka, O., Šimberová, I., Topolová, I. and Trenz, O.: *Měření podnikové výkonnosti.* Littera, Brno, 2013.
[9] Maple Online Help. 2014. http://www.maplesoft.com/support/help/ (last accessed 15.04.14.).
[10] Maplesoft. *Overview of the Optimization Package*. [online]. Maple Online Help. 2014. http://www.maplesoft.com/support/help/Maple/view.aspx?path=Optimization (last accessed 15.04.14.).
[11] Zaoralová, H.: *Lineární programování jako účinná zbraň v rukou manažera.* Fakulta podnikatelská Vysoké učení technické v Brně, Brno, 2012. Supervisor RNDr. Zuzana Chvátalová, Ph.D.

Analysis of Bank Efficiency: An Application of DEA Approach in the Czech Commercial Banks

Lucie Chytilová [1]

Abstract. In this article, we utilize Data Envelopment Analysis (DEA) approach to estimate the relative efficiencies of the Czech commercial banks from the year 2010 to the year 2013. Each bank is modeled as a linear system with some inputs and outputs. Inputs and outputs for this issue are selected with the help of intermediation approach that found Sealey and Lindley in [6]. Inputs are considered as taken client deposits, capital and personal costs per employee and outputs are denoted net loans and net interest income. All these inputs and outputs are extracted from annual reports of the Czech commercial banks which were during the period reviewed occurred at the territory of the Czech Republic. The main DEA models - CCR and BCC, are used. The results of these applications are paired with the results of other previous studies e.g. efficiency of banks in the V4 countries, Stavárek in [7] and the DEA theory. Similarly to the previous studies this article summarizes the different approaches for the issue and compares the results of all used models. Subsequently, there is selected the best model for the Czech banking industry for the period between years 2010 and 2013. To measure the relative efficiency of the commercial banks MS Excel and DEAFrontier software are used.

Keywords: Data Envelopment Analysis, Efficiency, Banking.

JEL classification: C6, G21.
AMS classification: 90B30, 90C08, 90C30, 90C90.

1 Introduction

The bank efficiency is very important and crucial issue especially in transition economies. There are numerous studies on the banking system efficiency for countries around the world. However, the East Europe is a unique because of its history. Banking sectors have faced ownership changes in the structure as a result of privatization and therefore changes in the legislative environment and institutional rules. There were also changes due to entry of new banks and the disappearance of old ones. Therefore, this article is focused on the banking system and the position of banks in the Czech Republic. There are analyzed banks which were in the bank industry during the period from 2010 to 2013 and their data were available.

Data envelopment analysis (DEA) is a non-parametric linear programming based technique for measuring the relative efficiency of a set of similar units (DMUs). Since the work of Charnes et al. in [3], DEA has demonstrated to be an effective technique for measuring the relative efficiency of a set of homogenous DMUs. In applications, DMUs may include banks, supermarkets, hospitals schools and other. The first two well-known DEA models are called CCR and BCC. The CCR model is formulated for constant returns to scale (CSR) situation. In [1] Banker at al. extended the origin model to variable returns to scale (VRS) situation. There are other types of DEA models such as additive or super efficiency model. DEA models are looking for an efficient frontier that envelops all data. These models are able to classify DMUs as efficient, located on the efficient frontier, and as inefficient, are not located on the efficient frontier.

A debatable topic in DEA is how to choose the appropriate inputs and outputs. There exist a lot of studies that focus on banking efficiency and the problem in the concept of inputs and outputs in the banking business. There are many different perspectives which describe the function of bank and express the essence of their activities. The used approached in this article is identified by Sealey and Lindley in

[1]VŠB - Technical University of Ostrava, Faculty of Economics, Sokolská 33, Ostrava, 701 21, Czech Republic, lucie.chytilova@vsb.cz

[6]. This approached is called as intermediation approach. It sees banks as institutions that transform money from deposits into funds provided to borrowers in the form of loans.

In [7] is utilized the CCR and BCC models. It showed that the best model for measuring the efficiency score of commercial banks in the East Europe is the input-oriented-BCC. This article verifies the result in the Czech bank industry for similar period of time. The obtain results by applying the BCC models were identify better then by applying the CCR. The input-oriented and output-oriented BCC models had the same numbers of efficient DMUs but the average value of mean efficiency are higher for input-oriented-BCC model.

2 Methodology of DEA

Data Envelopment Analysis (DEA) is a non-parametric approach for measuring the relative efficiency of decision making units (DMUs) when the data for multiple inputs and outputs are known exactly. Suppose that there is a set of T DMUs (DMU_k for $k = 1, ..., T$), let inputs and outputs data be $X = \{x_{ik}, i = 1, ...R; k = 1, ...T\}$ and $Y = \{y_{jk}, j = 1, ...S; k = 1, ...T\}$, respectively. Also, u_i for $i = 1, ...R$ and v_j for $j = 1, ..., S$ be the weights of i^{th} inputs and j^{th} outputs, respectively. Mathematically, the relative efficiency score of DMU_k can be defined as:

$$e_k = \frac{\sum_{j=1}^{S} v_j y_{jk}}{\sum_{i=1}^{R} u_i x_{ik}}, \text{ for } k = 1, ..., T. \tag{1}$$

In [2] Charnes et al. proposed the following CCR model to measure the efficiency score of the under evaluation unit, DMU_Q where $Q \in \{1, ..., T\}$:

$$\begin{array}{lll} & \max\ e_Q = \frac{\sum_{j=1}^{S} v_j \cdot y_{jQ}}{\sum_{i=1}^{R} u_i \cdot x_{iQ}}, & \\ \text{s.t.} & \sum_{j=1}^{S} v_j \cdot y_{jk} - \sum_{i=1}^{R} u_i \cdot x_{ik} \leq 0, & k = 1, ..., T, \\ & u_i \geq 0, & i = 1, ..., R, \\ & v_j \geq 0, & j = 1, ..., S. \end{array} \tag{2}$$

The model (2) is non-linear, more precisely it is linear-fractional programming. However, the model (2) could be easily transferred by Charmes-Cooper transformation to the standard linear programming problem. The transformed model (2) is:

$$\begin{array}{lll} & \max\ e_Q = \sum_{j=1}^{S} v_j \cdot y_{jQ}, & \\ \text{s.t.} & \sum_{i=1}^{R} u_i \cdot x_{iQ} = 1, & \\ & \sum_{j=1}^{S} v_j \cdot y_{jk} - \sum_{i=1}^{R} u_i \cdot x_{ik} \leq 0, & k = 1, ..., T, \\ & u_i \geq 0, & i = 1, ..., R, \\ & v_j \geq 0, & j = 1, ..., S, \end{array} \tag{3}$$

where $Q \in \{1, ..., T\}$

This model must be solved one for each DMU. Notice that in this model DMU_Q is CCR-efficient if and only if $e^* = 1$ and if there exists at least one optimal solution $(\mathbf{u}^*, \mathbf{v}^*)$ with $\mathbf{u}^* > \mathbf{0}$ and $\mathbf{v}^* > \mathbf{0}$ for the set $Q \in \{1, ..., T\}$. An inefficient units have a degree of relative efficiency that belongs to interval $[0, 1)$. The model (3) is called a multiplier form of input-orient-CCR model.

However, for computing and data interpretation is preferable to work with model that is dual associated to model (3). The model is referred as envelopment form of input-oriented CCR model:

$$\begin{array}{lll} & \min\ \theta_Q, & \\ \text{s.t.} & \sum_{k=1}^{T} \lambda_k \cdot x_{ik} - \theta_Q \cdot x_{iQ} \leq 0, & i = 1, ..., R, \\ & \sum_{k=1}^{T} \lambda_k \cdot y_{ik} - y_{iQ} \geq 0, & i = 1, ..., S, \\ & \lambda_k \geq 0, & k = 1, ..., T, \\ & \theta_Q \in (-\infty, \infty), & \end{array} \tag{4}$$

where λ_k is the weight for DMU_k for $k = 1, ..., T$. It is dual-variable unit. The θ_Q represents the efficiency score of DMU_Q. It can also be interpreted as a reduction rate of inputs to reach the efficient frontier.

There also exists the multiplier form of output-oriented CCR model. The output-oriented CCR model gives same results as the input-orient-CCR model. It can be seen in [2].

Banker et al. in [1] extended the CCR model. This model is called BCC model. The extension considers variable returns to scale assumption. This model has convex envelope of data which leads to more efficient DMUs. The mathematical model of dual multiplier form of input-oriented BCC model is as bellow:

$$\begin{aligned} & \max e_Q = \textstyle\sum_{j=1}^{S} v_j \cdot y_{jQ} - v_0, \\ \text{s.t.} \quad & \textstyle\sum_{i=1}^{R} u_i \cdot x_{iQ} = 1, \\ & \textstyle\sum_{j=1}^{S} v_j \cdot y_{jk} - \sum_{i=1}^{R} u_i \cdot x_{ik} - v_0 \leq 0, \quad k = 1, ..., T, \\ & u_i \geq 0, \qquad\qquad\qquad\qquad\quad i = 1, ..., R, \\ & v_j \geq 0, \qquad\qquad\qquad\qquad\quad j = 1, ..., S, \\ & v_0 \in (-\infty, \infty), \end{aligned} \tag{5}$$

where v_0 is the dual-variable assigned to the convexity condition $\mathbf{e^T}\lambda = \mathbf{1}$ of envelopment form of input-oriented BCC model.

The input-oriented-BCC model evaluates the efficiency of DMU_Q by solving the following envelopment form of linear program:

$$\begin{aligned} & \min e_Q = \theta_Q \\ \text{s.t.} \quad & \theta_Q \cdot \mathbf{x_Q} - \mathbf{X} \cdot \lambda \geq 0, \\ & \mathbf{Y} \cdot \lambda \geq \mathbf{y_Q}, \\ & \mathbf{e^T} \cdot \lambda = \mathbf{1}, \\ & \lambda \geq 0, \end{aligned} \tag{6}$$

where $e^T = (1, 1, , 1)$, $\theta_Q \in (-\infty, \infty)$ and λ is the vector of the dual variables units. DMU_Q is BCC-efficient the radial variable θ_Q is equal to one, ie. the optimal value of the objective function of the model (7) $e^* = 1$, otherwise it is BCC-inefficient. Units that are not effective have a value $e^* \in [0, 1)$.

The envelopment form of the output-oriented-BCC model is written as:

$$\begin{aligned} & \max e_Q = \eta_Q \\ \text{s.t.} \quad & \mathbf{X} \cdot \lambda \leq \mathbf{x_Q}, \\ & \eta_Q \cdot \mathbf{y_Q} - \mathbf{Y} \cdot \lambda \leq 0, \\ & \mathbf{e^T} \cdot \lambda = \mathbf{1}, \\ & \lambda \geq 0, \end{aligned} \tag{7}$$

The double multiplier form of the output-oriented-BCC model is associated with the above, is defined analogously as the relationship between inputs-oriented-BCC models and with similar properties. More details are in [3]. The optimal solutions of the output-oriented models are written as $g^* = \frac{1}{e^*}$, where e^* is the optimal solution for output-oriented model. It is for the better comparison of the optimal solutions.

3 Determination of inputs and outputs in banking industry

The literature that focuses on banking efficiency has a problem in the concept of inputs and outputs. There are many different perspectives on the determination of inputs and outputs. They try to describe the function of banks and express the essence of the activities. Some approaches perceive banks as producers of deposits and loans accounts, others as institutions that transforms money from deposits to funds in the form of loans. In this article, the second option is taken, so-called the *intermediation approach*. This approache was published by Sealey and Lindley in [6]. They have reported that the process of transformation of funds is due to multidimensional inconsistency between deposits and loans, and between creditors and debtors. Deposits are characterized as divisible, liquidity, and risk-free, loans are treated as indivisible, illiquid and risky. So inputs are usually defined as deposits and other sources of financial markets, together with interest and operating costs. The outputs are loans and other investments in yield interest income. The selected inputs and outputs for this article are written in Table 1.

inputs	outputs
taken client deposits (TD)	net loans (TL)
capita (FA)	net interest income (NII)
personnel costs per employee (PC)	

Table 1 The selected of inputs and outputs

Deposits (TD) are measured as the total amount of current and term deposits, which the bank received from individual clients and from other financial institutions. Capital (FA) is represented by the average book value of fixed assets. Labor is expressed by personnel costs (PC), which include wages and health and social insurance. Loans (TL) are defined as the net value of loans to customers and other financial institutions. Net interest income (NI) is the difference between total interest income and total costs.

Inputs and outputs for this article were obtained from annual reports of eighteen Czech banks. All data are in mil. CZK. The first analyzed period was from 2000 to 2013. During this time the Czech banking industry saw various developments - banks were connected, closed down or new banks entered the industry. The long time interval brought the problem to the retrieval, e.g. between 2000 and 2003 there were obtained data just only for 8 banks. 18 banks were analyzed only in 2012 and 2011. The rest is shown in Figure 1. The used data are on the link: http://mme2014.upol.cz/downloads/142-Appendix.pdf.

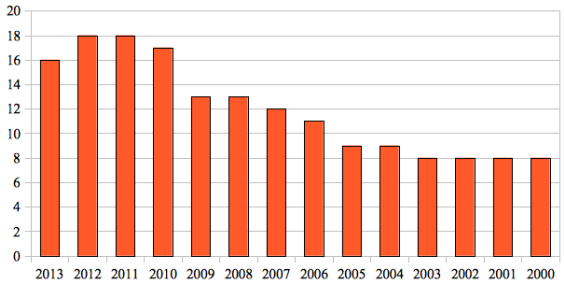


Figure 1 Number of analyzed bank in individual years

In order to discrimination power of the models are used just data from 2010 to 2013. There are analyzed 16 and more banks. This is due to the fact that the general rule for the least number of the evaluated units is that it must be at least three times higher then the number of inputs and output.

4 The estimation of the relative efficiency by DEA

Using the DEA non-parametric techniques have been established measure of relative efficiency for all banks analyzed separately for constant (CCR) and variable (BCC) returns to scale. The calculation is performed by using DEAFrontier software and MS Excel. Data of commercial banks in the Czech Republic are used to define a common efficient frontier. Primary elements of the descriptive statistics for calculated values of the relative efficiency of CCR and BCC input-oriented models are summarized in Table 2 and for output-oriented-BCC model are in Table 3. There are no values of maximum in these tables because all efficient units have value one and that is the maximum. Also in Table 3 there is no value of median. All values of median for output-oriented-BCC model are equal to one.

year	number of banks	number of efficient banks		mean		median		standard deviation		minimum	
		CCR	BCC	CCR	BCC	CCR	BCC	CCR	BCC	CCR	BCC
2010	17	6	11	0.722	0.831	0.672	1.000	0.196	0.184	0.531	0.532
2011	18	5	10	0.628	0.882	0.565	1.000	0.285	0.168	0.009	0.509
2012	18	5	13	0.617	0.871	0.559	1.000	0.297	0.248	0.117	0.139
2013	16	4	10	0.506	0.829	0.383	1.000	0.318	0.258	0.115	0.236

Table 2 Descriptive statistics of the relative efficiency of input-oriented models

Table 2 shows that level of efficiency with the VRS assumption achieves higher values than the efficiency with the CRS assumption. In other words, input-oriented-BCC models identify more banks as an efficient. These results are same for the whole period. According to the average mean value, the most efficient year was 2010 for the CCR model, the value is equal to 0.722. In this year was analyzed 17 banks and 6 were identified as CCR-efficient. According to the same criteria, the most efficient banks for input-oriented-BCC model are in 2011. There are analyzed 18 banks and 10 banks are BCC-efficient. In this model the value is around 0.8 for all the period. The years 2010 and 2013 have values lower then the rest years. In 2010 this could be cost by the legacy of the economical crisis and global problem of banks which. In 2013 this could be cost by limit value of banks in this analysis. However, similar conclusions for bank industry were found earlier, for example in [7] by Stavárek.

year	number of number of banks	number of efficient banks	mean	standard standard deviation	minimum
2010	17	11	0.841	0.177	0.532
2011	18	10	0.830	0.280	0.014
2012	18	13	0.860	0.280	0.133
2013	16	10	0.808	0.302	0.138

Table 3 Descriptive statistics of the relative efficiency of output-oriented-BCC models

Table 3 shows the descriptive statistics just for output-oriented-BCC model. Values of output-oriented-BCC model are similar to input-oriented-BCC. However, in general it could be said that the mean values of output-oriented-BCC model are lower then for input-oriented-BCC. These models have same numbers of efficient banks. On other hand, these models do not identified the same years as the most efficient. According to the average mean value, the most efficient year for this model was the year 2012, the value is equal to 0.860. In this year was analyzed 18 banks and 13 were identified as BCC-efficient. The lowest average mean value is in the year 2013, this value is equal to 0.808. The year 2013 has the lowest mean value both BCC models.

Table 4 shows which banks are included in the analysis. It can also be seen which years are known. Table 4 shows which bank is efficient and in which year is this bank efficient.

name of annualyzed banks	available data	2013	2012	2011	2010
Air Bank a.s.	2013-10				**Y**
Česká exportní banka	2013-10	**Y**	**Y**	**Y**	**Y**
Česká spořitelna	2013-10	O	O	O	O
ČMZRB	2013-10	**Y**	O		
ČSOB	2013-10	O	O	O	O
Equa bank	2013-11				
Evropsko-ruská banka	2013-10		O	O	
Fio banka, a.s.	2013-10	**Y**	**Y**		**Y**
GE Money Bank	2013-10	O	**Y**	**Y**	**Y**
J&T Banka	2013-10		O		
Komerční banka	2013-10	O	O	O	O
LBBW Bank	2012-10				
Poštová banka	2013-10	O	**Y**	**Y**	
PPF banka	2013-10	O	O	**Y**	**Y**
Raifeisenbank	2013-10				O
Sberbank CZ	2013-10				
UniCredit Bank	2013-10	O	O	O	O
ZUNO BANG AG	2012-10		**Y**	**Y**	**Y**

Table 4 Efficient banks in different years

Markings "**Y**" and "O" mean that the bank is efficient for the year. If the marking is "**Y**", then the bank is CCR-efficient and BCC-efficient. If the marking is "O", then the bank is just BCC-efficient. As already stated, input and output-oriented-BCC models identify same banks as efficient so there is no need to create different marking. For example, from Table 4 may be seen that *PPF banka* has data for all period. Also it is seen that this bank is BCC-efficient for the all period but CCR-efficient is only for the period from 2013 to 2012. Table 4 also proves again that there are more bank BCC-efficient than CCR-efficient.

5 Conclusion

In the article is estimated and analyzed the efficiency of the Czech banking system for years from 2010 to 2013. There are used inputs and outputs indicators of the intermediation approach for banking efficiency, namely taken client deposits, capital and personal costs per employee as inputs and net loans, net interest income as outputs. In order to calculate the bank efficiency scores' measurement there are used the DEA approaches. The data are taken from the annual reports of the available banks.

There are used two basic DEA models for the analyses of the Czech banking system. The first used model is CCR model and the second is the BCC model. The BCC model has two modifications - input and output-orientation, and both are used. The article illustrates that if there is used the intermediation approach of the selection of the set of inputs and outputs. And there are used the basic DEA models then the efficiency for bank business in the Czech Republic is best for BCC model. That means that BCC models give results with higher efficient DNUs than CCB models. Specifically, the best model is the input-oriented-BCC model. It is same as in [7] by Stavárek. This has been proved through the mean values of efficient banks for the years. These values of relative efficient banks are higher for BBC models than for CCR. There is just small difference in these values between input and output-oriented-BCC models, but still higher values of efficient baks are known for output-oriented-BCC.

In future should be models used for larger database of data or for date in different region. It would be good too to try also new types of DEA models e.g. additive model - SBM or fuzzy DEA models and other. Also how it was said there exist problem with the approaches for choosing the right inputs and outputs for the banking business. So the different types of DEA models could help to test these approaches of choosing the set of inputs and outputs and find out the best for DEA models.

Acknowledgements

The research is due to the support provided by the project SGS VSB-TUO SP2014/146. All support is greatly acknowledged.

References

[1] Banker, R.D., Charnes, A., and Cooper, W.W.: Some models for estimating technical and scale inefficiency in data envelopment analysis, *Management Science* **30** (1984), 1078-1092.

[2] Charnes, A., Cooper, W.W., and Rhode, E.: Measuring the efficiency of decision making units, *European Journal of Operational Research* **2** (1978), 429-444.

[3] Cooper, W. W., Seiford, M.L., and Tone, K.: *Data Envelopment Analysis*. Springer, New York, 2007.

[4] Fiala, P. *Operační výzkum - Nové trendy*. Professional Publishing, Praha, 2010.

[5] Jablonský, J., and Dlouhý, M.: *Modely hodnocení efektivnosti produkčních jednotek*. Professional Publishing, Praha, 2010.

[6] Sealy, C. W. Jr., and Lindley J. T.: Inputs, Outputs and a Theory of Production and Cost at Depository Financial Institutions, *Journal of Finance* **32** (1977), 1251-1266.

[7] Stavárek, D.: *Restrukturalizace bankovních sektorú a efektivnosti bank v zemích Visegrádské skupiny*. Studia Oeconomica, Karviná, 2005.

Open Vehicle Routing Problem

Zuzana Čičková[1], Ivan Brezina[2], Juraj Pekár[3]

Abstract. Routing and scheduling of vehicles are important operational issues in the logistics management. This paper deals with the open vehicle routing problem (OVRP). The OVRP is an extension of classical capacitated vehicle routing problem (CVRP) that is well known in optimization. The goal of the OVRP is to find optimal shortest route for a vehicle (vehicles) with limited capacity in order to serve a given set of customers. The difference between the OVRP and the classical CVRP model is that the vehicle (vehicles) does (do) not need to return to the depot after the service of the last customer. In this paper, we formulate the mathematical models to capture all aspects of that problem based on mixed integer programming (MIP) with linear objective function and constraints. The experiments were provided on the free published instance for CVRP.

Keywords: Open Vehicle Routing Problem, Mixed Integer Programming, Mathematical Model.

JEL Classification: C02, C61
AMS Classification: 90C11, 90B06

1 Introduction

Logistics is evidently one of the major dimensions of many firms and the related logistics costs constitute a significant share of the total costs of every organization. Many variants of routing and scheduling problems that can be very rewarding are known in the field of logistics. The well-known capacitated vehicle routing problem (CVRP) is one of the most discussed problems in operations research [1], [2], [3], [4], [6], [7], [8]. The importance of that problem is evidently seen in the field of operations research or artificial intelligence etc., because of its computational complexity, but also the importance follows from a great practical applicability, thus the problem can be applied in more general way.

This paper describes the open vehicle routing problem (OVRP). The OVRP is a popular problem in the field of distribution management and can be used to model many real-life problems. For example if the company deals without its own vehicle fleet and has to hire some vehicle to deliver its products to customers. In this case, the company is not concerned whether the vehicle returns to the depot, and does not pay the traveling costs between the last customer and the depot. The OVRP can be described as follows: Consider a depot from which some products have to be delivered to a set of customers. Products are loaded on a vehicle (vehicles) at the depot and afterwards load needs to be transported to the customers. The capacity of vehicle (or fleet of vehicles) is (are) known (if we use more than one vehicle, we suppose the same capacity of all of them), so that all customers' demand needs to be served with the use of some vehicle (all the demands are met in full). We assume known shortest distance between depot and each customer's location, as well as between each pairs of customer's location. The goal is to find so optimal shortest route for a vehicle (vehicles) so that each customer demand is met.

2 Open vehicle routing problem

The OVRP can be stated as follows: Consider graph $G = (N_0, A)$, where $N_0 = \{0, 1,\ldots n\}$ the set of all nodes in the graph, so that $N_0 = N \cup \{0\}$, where the set $N = \{1,\ 2...\ n\}$ represents the set of served nodes (customers) and the node indexed 0 represents origin (depot). The set $A = \{(i, j)\colon i, j \in N_0,\ i \neq j\}$ is the arc set of G. A shortes distance time d_{ij} is associated with every arc of the graph. Let the parameters $q_i,\ i \in N$ represent demand of customer and parameter g represents the capacity of vehicle (vehicles). Mathematical programming formulation of OVRP requires two type of variables: the binary variables $x_{ij},\ i, j \in N_0$ with a following notation: $x_{ij} = 1$ if customer i precedes customer j in a route of the vehicle and $x_{ij} = 0$ otherwise. Further on, we will apply the variables $u_i,\ i \in N$ that based on well-known Miller –Tucker-Zemlin's formulation [5] of the traveling salesman problem. That variables will represent cumulative demand of customers on one particular route.

Further on, we distinguish two alternatives:

[1] University of Economics in Bratislava, zuzana.cickova@euba.sk.
[2] University of Economics in Bratislava, ivan.brezina@euba.sk.
[3] University of Economics in Bratislava, juraj.pekar@euba.sk.

1. Let consider a fllet of vehicles (we assume unlimited number of vehicles), if the vehicle is not required it will be returned to the depot ($OVRP_m$)
2. Let consider only one vehicle, which held a number of routes and after servicing the last of customer , the vehicle is not required to be returned to depot ($OVRP_o$)

In the first case ($OVRP_m$) we can state mathematical model as follows:

$$\min f(\mathbf{X},\mathbf{u}) = \sum_{i\in N_0}\sum_{j\in N_0} d_{ij}x_{ij} \tag{1}$$

$$\sum_{i\in N_0} x_{ij} = 1,\ j\in N,\ i\neq j \tag{2}$$

$$\sum_{j\in N_0} x_{ij} \leq 1,\ i\in N,\ i\neq j \tag{3}$$

$$\sum_{j\in N} x_{0j} \geq 1 \tag{4}$$

$$u_i + q_j - g(1-x_{ij}) \leq u_j,\ i\in N_0,\ j\in N,\ i\neq j \tag{5}$$

$$q_i \leq u_i \leq g,\ i\in N \tag{6}$$

$$u_0 = 0 \tag{7}$$

$$x_{ij}\in\{0,\ 1\},\ i,j\in N_0,\ i\neq j \tag{8}$$

The objective function (1) models the total distance travelled by all of the vehicles. Equations (2) ensure that only one of the vehicles enters each customer exantly once and equations (3) ensure that the corresponding vehicle does not need to depart from every customer, because the route ends after serving the last of them. Equation (4) ensure that the each of the vehicles starts its route exactly once. Equations (5) avoid the presence of sub-tour and and also calculates the real cumulative demads of customers for the next node on the route based on previous node. Equations (6) ensure that all demands on the route must not exceed the capacity of the vehicle. The fix values of variables u_0 are set up by equation (7).

Consider now the use of only one vehicle that can take multiple routes, and ends its route at the last-served customer. In the mathematical formulation one can use a fictitious n + 1-node and consider following: $d_{i,n+1} = 0$ and $d_{n+1,j} = 0$ and $q_{n+1} = 1$. Let N_2 and N_3 be the sets: $N_2 = N_0 \cup \{n+1\}$ a $N_3 = N \cup \{n+1\}$. Then, the mathematical model ($OVRP_o$) can be formulated as follows:

$$\min f(\mathbf{X},\mathbf{u}) = \sum_{i\in N_2}\sum_{j\in N_2} d_{ij}x_{ij} \tag{9}$$

$$\sum_{i\in N_2} x_{ij} = 1,\ j\in N_3,\ i\neq j \tag{10}$$

$$\sum_{j\in N_2} x_{ij} = 1,\ i\in N_3,\ i\neq j \tag{11}$$

$$u_i + q_j - g(1-x_{ij}) \leq u_j,\ i\in N_2,\ j\in N_3,\ i\neq j \tag{12}$$

$$q_i \leq u_i \leq g,\ i\in N \tag{13}$$

$$u_0 = 0 \tag{14}$$

$$x_{ij}\in\{0,\ 1\},\ i,j\in N_0,\ i\neq j \tag{15}$$

$$x_{n+1,0} = 1 \tag{16}$$

$$x_{0,n+1} = 0 \tag{17}$$

The objective function (9) models the total distance of vehicle route. Equations (10) ensure that the vehicle enters to every customer exactly once and constraints (11) ensure that the vehicle leaves every customer exactly once (including fictional node). Equations (12) avoid the presence of sub-tour and and also calculates the real cumulative

demads of customers for the next node on the route based on previous node. Equations (13) ensure that all demands on the route must not exceed the capacity of the vehicle. The fix values of variables are set up by equations (14), (16) and (17). The equation (17) also ensure the fact that the vehicle is not required to be returned to depot after servicing the last of customers (it is returned from the fictional node).

3 Computational experiments

The computational experiments were provided on the base of free available benchmark instances of CVRP[4]. The mathematical formulations of $OVRP_m$ and $OVRP_o$ were implemented in software GAMS (solver Cplex 12.2.0.0) on PC INTEL(R) Core(TM) 2 CPU, E8500 @ 3.16 GHz, 3.25 GB RAM under Windows XP. The results are shown in Table 1.

		CVRP		$OVRP_o$		$OVRP_m$	
Instance	Dimension	Value of objective function	Time (s)	Value of objective function	Time (s)	Value of objective function	Time (s)
rc206_1	4	153,22	0,01	110,2086	0,01	76,67	0,02
rc207_4	6	87,04	0,13	66,4244	0,03	51,66	0,01
rc205_1	14	163,28	208,59	136,49	479,64	109,03	0,2
rc202_2	14	211,91	92,25	255,78	575,55	170,73	1,41
rc203_4	15	220,44	36,73	182,36	3,06	144,28	0,03
r101_25	26	335,26	1365,81	311,07	67,75	291,51	0,03
E-n13-k4	13	247	37,44	216	22	150	0,28
E-n22-k4	22	375	76528,23	335,56	8407,11	255,03	0,13
E-n23-k3	23	568,55	734,88	507,82	124,08	431,19	0,41
Eil7	8	132	0,2	102	0,25	73	0,01

Table 1 Solution of some benchmark instances

The Table 1 compares the solution of classical CVRP to its open versions (presented models of $OVRP_o$ and $OVRP_m$). The first column of the table show the name of instance and second column shows the number of nodes in instance (including depot). For each instance we report the solution of CVRP, $OVRP_o$ and $OVRP_m$ (first column shows the value of objective function and the next column gives number of seconds used for finding the solutions). In table we could already see that the needed computing time for finding solution of $OVRP_m$ is significantly shorter compared to $OVRP_m$ (this fact is implied by different structure of presented models).

4 Conclusion

This paper considers a modification of capacited vehicle routing problem named open vehicle vehicle routing problem, which addresses practical issues in routing vehicle (or fleet of vehicles) in case that vehicle (vehicles) does (do) not need to return to the depot after a service of the last customer. The problems were formulated on the base of mixed integer programming (MIP) with linear objective function and constraints. So that formulation allows the use of standard software for solving MIP problems. The computational experiments used software GAMS based on CVRP instances are presented. The comparision to open versions of CVRP is based on value of objective functions as well as on number of seconds used for finding the solutions.

[4] http://myweb.uiowa.edu/bthoa/TSPTWbenchmarkDataSets.htm
http://www.coin-or.org/SYMPHONY/branchandcut/VRP/data/index.htm#E
http://w.cba.neu.edu/~msolomon/problems.htm
http://people.brunel.ac.uk/~mastjjb/jeb/orlib/files/

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic – VEGA, grant no. 1/0104/12 „Modeling supply chain pricing policy in a competitive environment“.

References

[1] Čičková, Z., Brezina, I.: An evolutionary approach for solving vehicle routing problem. In: *Quantitative methods in economics multiple criteria decision making XIV*. IURA EDITION, Bratislava, (2008), 40–44.

[2] Čičková, Z., Brezina, I., Pekár, J.: *Solving the Real-life Vehice Routing Problem with Time Windows Using Self Organizing Migrating Algorithm*. Ekonomický časopis **61**(5/2013), 497–513.

[3] Desrochers, M., Desrosiers, J., Solomon, M.: A New Algorithm for the Vehicle Routing Problem with Time Windows. *Operations Research*. **40** (1992), 342-354.

[4] Fábry, J., Kořenář, V., Kobzareva, M.: Column Generation Method for the Vehicle Routing Problem – The Case Study. Janská Dolina 06.09.2011 – 09.09.2011. In: *Mathematical Methods in Economics 2011 [CD-ROM]*. Praha, PROFESSIONAL PUBLISHING, (2011), 140–144.

[5] Miller, C.E., Tucker, A.W., Yemlin, R.A.: *Integer programming Formulation of Traveling Salesman Problems*. Journal of the ACM, 7(4/1960), 326-329.

[6] Palúch, S., Peško, Š.: *Kvantitatívne metódy v logistike*. EDIS vydavateľstvo ŽU, Žilina, (2006).

[7] Peško, Š.: *Optimalizácia rozvozov s časovými oknami.* Konferencia KYBERNETIKA - história, perspektívy, teória a prax, zborník referátov, Žilina, (2002), 76–81.

[8] Yu, B., Yang, Z., YAO, B.: An improved ant colony optimization for vehicle routing problem. *European Journal of Operational Research* **1** (2009), 171–176.

Split Delivery Vehicle Routing Problem with Time Windows

Zuzana Čičková[1], Marian Reiff[2], Kvetoslava Surmanová[3]

Abstract. Vehicle routing problem is well known problem in the area of logistics distribution. In classical capacitated vehicle routing problem formulation it is assumed, that customers demand is smaller than capacity of vehicle and each customer order is met in full. Last assumption can cause exclusion of customer from vehicle route for the capacity exceeded reason. In case of split deliveries it is possible to visit customer more than once, in different vehicle route and that can result in transportation cost reduction. This kind of problem is known as split delivery vehicle routing problem. Another frequent extension of vehicle routing problem is incorporation of customer service time requirement. In this paper, we formulate a mathematical model to capture all aspects of that problem based on mixed integer programming (MIP) with linear objective function and constraints.

Keywords: Vehicle Routing Problem, Split Delivery, Time Windows.

JEL Classification: C02, C61
AMS Classification: 90C11, 90B06

1 Introduction

The presented paper describes the split delivery vehicle routing problem (SDVRP). Split delivery vehicle routing problem is modification of well-known classical capacitated vehicle routing problem [2], [7]. Different types of modification of classical capacitated vehicle routing problem (CVRP) have attracted the interest of various researchers in the field of management science and can help to model many real-life distribution problems and situations. In classical capacitated vehicle routing problem, aims to find a set of routes at a minimal cost (finding the shortest path) beginning and ending the route at the depot, so that the known demand of all nodes are fulfilled. It is assumed, that customers demand is smaller than capacity of vehicle and each customer order is met in full. Last assumption can cause customer exclusion from vehicle route for vehicle capacity exceeded reason. In case of split deliveries it is possible to visit customer more than once, in different vehicle route and that can result in transportation cost reduction. This kind of problem is known as split delivery vehicle routing problem. SDVRP is a relaxation of the classical capacitated vehicle routing problem. The problem was first published by Dror and Trudeau in [5], who empirically showed that allowing split deliveries can significantly reduce the total distance travelled. Archetti et al. [1] prove that this reduction can achieve up to 50%. The practical problems of physical distribution often include the need to respect the time restriction. Mathematical formulation SDVRP based on integer programming was introduced by Dror et al. in [6] Frequently we consider time restriction that are a consequence of earliest possible time of service, the latest acceptable time of service or the need to serve during the given time interval. The above mentioned terms are known as time windows. So we can talk about the split delivery vehicle routing problem with time windows (SDVRPTW). Frizzell and Giffin in [8] are among the first who considered time windows constraints in SDVRP followed by [3], [9], [12]. If it is necessary to consider only the earliest possible time of service or the last acceptable time of service, the problem is known as problem with soft delivery time windows, if we are dealing with time interval with given lower and upper limit, those problems are known as problems with hard delivery time windows, e. g. [4]. Another description of soft windows can be found e.g. in [11], where the violation of time restriction is allowed, although incurring some cost. In formulation of that problem we consider that the shortest travel time between depot and each customer's location is known, as well as between each pairs of customer's location shortest travel time is known.

2 Split Delivery Vehicle Routing Problem

The mathematical model of SDVRP can be describeded on graph $G = (N_0, A)$, where set $N_0 = \{0, 1,\ldots n\}$ represents the set of all nodes, so that $N_0 = N \cup \{0\}$ and where the set $N = \{1, 2... n\}$ represents the set of served nodes (customers) The node represented by number 0 denotes origin (depot). The set $A = \{(i, j): i, j \in N_0, i \neq j\}$ represents the set of arcs on graph G. The shortes distance time d_{ij} is associated with every arc on the graph. Let the set $H = \{1, 2, \ldots h\}$ represents the fleet of vehicles with same capacity and with maximum number of h vehicles that can be used for distribution. We

[1] University of Economics in Bratislava, zuzana.cickova@euba.sk.
[2] University of Economics in Bratislava, marian.reiff@euba.sk.
[3] University of Economics in Bratislava, kvetoslava.surmanova@euba.sk.

will also consider the possibility that some of the vehicles may remain unused in the depot. Let the parameters q_i, $i \in N$ represent demand of customer and parameter g represents the capacity of vehicle (vehicles). Mathematical programming formulation of SDVRP requires two type of variables: the binary variables x_{ijh}, $i, j \in N_0$, $h \in H$ with a following notation: $x_{ijh} = 1$ if customer i precedes customer j in a route of the h-th vehicle and $x_{ijh} = 0$ otherwise. Further on, we will apply the variables u_{ih}, $i \in N$, $h \in H$ that represent cumulative demand of customers on the route of h-th vehicle and positive variables a_{ih} $i \in N$, $h \in H$ which represents part of i-th customer requirement, that is met with h-th vehicle. The goal is to set up the shortest vehicles routes so that all customers demands are met while not exceeding capacity of each vehicle.

The mathematical model can be stated as follows:

$$\min f(\mathbf{X}, \mathbf{U}, \mathbf{A}) = \sum_{i \in N_0} \sum_{j \in N_0} \sum_{h \in H} d_{ij} x_{ijh} \tag{1}$$

$$\sum_{j \in N} x_{0jh} \leq 1, \; h \in H \tag{2}$$

$$\sum_{i \in N_0} x_{ijh} = \sum_{i \in N_0} x_{jih}, \; j \in N, \; h \in H, \; i \neq j \tag{3}$$

$$u_{ih} + a_{ih} - g(1 - x_{ijh}) \leq u_{jh}, \; i \in N_0, \; j \in N, \; h \in H, i \neq j \tag{4}$$

$$a_{ih} \leq u_{ih} \leq g, \; i \in N, \; h \in H \tag{5}$$

$$\sum_{h=1}^{k} a_{ih} = q_i, \; i \in N \tag{6}$$

$$a_{ih} \leq q_i \sum_{j=1}^{n} x_{ijh}, \; i \in N, \; h \in H, \; i \neq j \tag{7}$$

$$u_{0h} = 0, \; h \in H \tag{8}$$

$$x_{ijh} \in \{0, 1\}, \; i, j \in N_0, \; h \in H, \; i \neq j \tag{9}$$

$$a_{ih} \geq 0, \; i \in N_0, \; h \in H \tag{10}$$

The objective function (1) models the total distance travelled by all of the vehicles. Equations (2) ensure each vehicle leaves the depot no more than once (i.e. some vehicles may remain unused in the depot). Constrain (3) ensures the continuity (contiguity) of routes for each vehicle. Equations (4) avoid the presence of sub-tour and also calculates the real cumulative demand of customers for the next node on the route based on previous node on the route of given vehicle. Together with equations (7) equations (4) summarize customer demands quantities realized by h-th vehicle. Requirement to meet whole customer's demand is ensured by the equations (6). The fix values of variables u_{0h}, $h \in H$ are set up by equations (8).

3 Split Delivery Vehicle Routing Problem with Time Windows

Let us take into account the before mentioned assumption and further on, let consider the time window for each customer, so that there is known given time interval, where e_i, $i \in N$ represents the first possible time to serve i-th customer and l_i, $i \in N$ represents the last acceptable time to leave the i-th customer. Also, there is known a service time o_i, $i \in N$ of each customer. In case the vehicle can´t provide service from time constraint reasons, vehicle will wait till earliest serving time. In addition to known parameters d_{ij}, which are associated with every arc of the graph, also the parameters t_{ij} that represent the travel time between node i and node j are known, $i, j \in N_0$. The goal is to set up the shortest vehicles routes so that all customers' demands are met by not exceeding capacity of each vehicle and serving customers in given time period. Variables τ_{ih}, $i \in N$, $h \in H$ that represent real starting time of service of i-th customer by h-th vehicle, are used instead of variable u_{ih}.

Thus, the mathematical models can be stated as follows:

$$\min f(\mathbf{X},\mathbf{A},\mathbf{T}) = \sum_{i\in N_0}\sum_{j\in N_0}\sum_{h\in H} d_{ij}x_{ijh} + v\sum_{i\in N_0}\sum_{j\in N_0}\sum_{h\in H} t_{ij}x_{ijh} \tag{11}$$

$$\sum_{j\in N} x_{0jh} \le 1,\ h\in H \tag{12}$$

$$\sum_{i\in N_0} x_{ijh} = \sum_{i\in N_0} x_{jih},\ j\in N,\ h\in H,\ i\ne j \tag{13}$$

$$\tau_{ih} + o_i + t_{ij} - M(1-x_{ijh}) \le \tau_{jh},\ i\in N_0,\ j\in N,\ h\in H, i\ne j \tag{14}$$

$$e_i \le \tau_{ih},\ i\in N,\ h\in H \tag{15}$$

$$\tau_{ih} + o_i \le l_i,\ i\in N,\ h\in H \tag{16}$$

$$\sum_{h=1}^{k} a_{ih} = q_i,\ i\in N \tag{17}$$

$$\sum_{i\in N} a_{ih} \le g,\ h\in H \tag{18}$$

$$a_{ih} \le q_i \sum_{j=1}^{n} x_{ijh},\ i\in N,\ h\in H,\ i\ne j \tag{19}$$

$$\tau_{0h} = 0,\ h\in H \tag{20}$$

$$x_{ijh} \in \{0,\ 1\},\ i,j\in N_0,\ h\in H,\ i\ne j \tag{21}$$

$$a_{ih} \ge 0,\ i\in N_0,\ h\in H \tag{22}$$

The objective function (11) models the total distance travelled by all of the vehicles and total time travelled by vehicles. Parameter v allows to set importance of the time criterion compared to distance criterion. Equations (12) ensure each vehicle leaves the depot no more than once (i.e. some vehicles may remain unused in the depot). Constrain (13) ensures the contiguity of routes for each vehicle. Let M represents large positive number, than equations (14) avoid the presence of sub-tour and also provide calculations of real customer starting time of service by h-th vehicle. Model equations (15) and (16) are ensuring time windows conditions. Requirement to meet whole customer's demand is ensured by the equations (17). Equations (18) provide not exceeding the capacity of each vehicle. Equations (19) summarize customer demands quantities realized by h-th vehicle. The fix values of variables τ_{0h}, $h\in H$ are set up by equations (20).

4 Empirical results

The principle of split delivery vehicle routing problem and split delivery vehicle routing problem with time windows are illustrated on data listed in Table 1, 2 and 3. Parameter g is set to value equal to 10 and parameter v to value 0,1. Also solution of classical CVRP is shown to illustrate difference between CVRP and SDVRP. The calculations were done with help of software GAMS (solver Cplex 12.2.0.0) on PC INTEL(R) Core(TM) 2 CPU, E8500 @ 3.16 GHz, 3.25 GB RAM under Windows XP.

i	1	2	3	4
q_i	8	3	4	4
e_i	1	2	10	4
l_i	5	15	15	5
o_i	0,3	0,5	0,8	0,5

Table 1 Initial parameters q_i, e_i, l_i, o_i

d_{ij}	0	1	2	3	4
0	0	200	115	123	161
1	200	0	85	121	48
2	115	85	0	84	46
3	123	121	84	0	73
4	161	48	46	73	0

Table 2 Shortest distance matrix

t_{ij}	0	1	2	3	4
0	0	3,4	1,4	1,5	3
1	3,4	0	1,4	2	1
2	2	1	0	1	1
3	2	2	1	0	0,7
4	4	0,8	0,5	0,8	0

Table 3 Shortest time matrix

The results are shown in following tables:

CVRP	Route	Distance
Vehicle 1	0 1 0	400
Demand realized	8	
Vehicle 2	0 3 0	246
Demand realized	4	
Vehicle 3	0 4 2 0	322
Demand realized	4 3	

Table 4 Classical capacitated vehicle routing problem solution

SDVRP	Route	Distance
Vehicle 1	0 2 1 0	400
Demand realized	2 8	
Vehicle 2	0 2 4 3 0	357
Demand realized	1 4 4	

Table 5 Split delivery vehicle routing problem solution

SDVRPTW	Route	Distance
Vehicle 1	0 2 4 3 0	357
Demand realized	3 4 3	
Staring time of service	2 4 10	
Waiting time	0,6 0,7 4,7	
Finishing time of service	2,3 4,5 10,5	
Vehicle 2	0 1 3 0	444
Demand realized	8 1	
Staring time of service	3,4 10	
Waiting time	0 4,3	
Finishing time of service	3,7 10,8	

Table 6 Split delivery vehicle routing problem with time windows solution

The total distance traveled for presented problems is shown and compared in table 7.

Total distance traveled	
CVRP	968
SDVRP	757
SDVRPTW	801

Table 7 Comparison of total distance traveled

5 Conclusion

This paper considers a modification of classical capacitated vehicle routing problem named split delivery vehicle routing problem and its further extension named split delivery vehicle routing problem with time windows, which addresses practical issues in routing problem if it is possible to visit customer more than once and if specific time period of service is required. The problems were formulated on the base of mixed integer programming (MIP) with linear objective function and constraints. So that formulation allows the use of standard software for solving MIP problems. Based on empirical result possible savings are presented and illustrated.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic – VEGA, grant no. 1/0104/12 „Modeling supply chain pricing policy in a competitive environment“.

References

[1] Archetti, C. Savelsbergh, M.W.P., Speranza, M.G.: Worst-case analysis for Split Delivery Vehicle Routing Problems. *Transportation Science* **40** (2006), 226-234.

[2] Čičková, Z., Brezina, I.: An evolutionary approach for solving vehicle routing problem. In: *Quantitative methods in economics multiple criteria decision making XIV*. IURA EDITION, Bratislava, 2008, 40–44.

[3] Čičková, Z., Brezina, I., Pekár, J.: Solving the Real-life Vehice Routing Problem with Time Windows Using Self Organizing Migrating Algorithm. *Ekonomický časopis* **61** (2013), 497–513.

[4] Desrosiers, J., Y., Dumas, Y., Solomon, M. M., Sournis, E.: Time Constrained Routing and Scheduling. *Handbooks in Operations Research and Management Science*, vol. 8, Network routing, chapter 2, 1995.

[5] Dror, M. Trudeau, P.: Savings by Split Delivery Routing. *Transportation Science* **23** (1989), 141-145.

[6] Dror, M., Laporte, G., Trudeau, P.: Vehicle routing with split deliveries. *Discrete and Applied Mathematics* **50** (1994), 239–254.

[7] Fábry, J., Kořenář, V., Kobzareva, M.: Column Generation Method for the Vehicle Routing Problem – The Case Study. In: *Proceedings of the Mathematical Methods in Economics Janská Dolina 06.09.2011 – 09.09.2011*. PROFESSIONAL PUBLISHING, Praha, 2011, 140–144.

[8] Frizzell, P.W., Giffin, J.W.: The split delivery vehicle scheduling problem with time windows and grid network distances. *Computers & Operations Research* **6** (1995), 655-667.

[9] Gendreau, M., Dejax, P., Feillet, D., Gueguen, C.: *Vehicle Routing with Time Windows and Split Deliveries, Technical Report.* Laboratoire d'Informatique d'Avignon, 2006.

[10] Pelikán, J., Fábry, J.: Split delivery vehicle routing problem. In: *Proceedings of the Mathematical Methods in Economics Kostelec nad Černými lesy 09.09.2009 – 11.09.2009.* Czech University of Life Sciences in Prague, Praha, 2009, 259–262.

[11] Peško, Š.: Optimalizácia rozvozov s časovými oknami. *In:* Proceedings of the *KYBERNETIKA - história, perspektívy, teória a prax,* Žilina, 2002, 76–81.

Forecasting unemployment in the Eurozone

Ondřej Čížek[1]

Abstract. The goal of the paper is to present a forecasting model of the unemployment rate. The model is applied in the context of the Eurozone. Parameter instability is considered due to the effects of the current economic crisis. The forecasting power of the model is assessed by generating in-sample forecasts. Finally, one-year-ahead out-of-sample forecasts are also performed and economically interpreted.

Keywords: labor market, unemployment rate, forecasting, in-sample, out-of-sample.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Unemployment is important economic variable indicating the state of the economy. Currently, high unemployment is the main problem of the Eurozone. The goal of this paper is formulation as well as application of the forecasting model of unemployment in the Eurozone. Chapter 2 presents dynamic linear model of unemployment. Although the model is formulated in continuous time, data are assumed to be available only at discrete dates. Data used for econometric estimation are described in chapter 3. Econometric estimation of the parameters of the model is the subject of chapter 4. Instability of the parameters due to current economic crisis are also discussed and statistically tested in this chapter. Chapter 5 deals with forecasting, which is the main subject of the paper. In-sample as well as out-of-sample forecasts are generated and economically interpreted. Final chapter 6 summarizes the main findings.

2 Dynamic model of unemployment

Following Shimer [6], the model of the labor market is formulated in continuous time environment in which data are available only at discrete dates. The model is based on the following assumptions:

- workers are only either employed or unemployed, they cannot enter or leave the workforce which is assumed to be constant over time;
- probability of finding (or loosing) a job is the same for all of the workers and these probabilities are assumed to be constant over time.

The assumption of constant probabilities over time is relaxed in Shimer [6], where the problem of measuring these probabilities is discussed. There is empirical evidence (Shimer [7]) according to which the rest of these assumptions are good approximation of reality. So the model based on these premises could be a reasonable approximation. These assumptions are nonetheless very strong. For this reason, the assumption of constant probabilities over time will be relaxed later in the next chapters. The assumption of constant labor force is relaxed in Barnichon, Nekarda [3].

The dynamics of the number of unemployed workers is the following:

$$\dot{U}(t) = s \cdot E(t) - f \cdot U(t), \tag{1}$$

where $U(t)$ is the number of unemployed workers at time instant t,

$E(t)$ is the number of employed workers at time instant t,

$\dot{U}(t) \equiv dU(t)/dt$

s is the job separation rate[2],

f is the job finding rate.

[1] University of Economics Prague, 4, W. Churchill Sq., CZ – 130 67, Prague 3, Faculty of Informatics and Statistics, Department of Econometrics, cizeko@vse.cz.
[2] Job separation rate is the mean value of how many times a (representative) worker looses a job during a period (one month).

This equation represents a Poisson process. The time length of the unemployment X is therefore a random variable with exponential probability distribution. The probability of being unemployed longer than one period (month) is thus given by

$$P(X \leq 1) = 1 - e^{-f \cdot 1}. \tag{2}$$

This probability will be called job finding probability and denoted by F. The probability of loosing a job will be denoted S and therefore:

$$F = 1 - e^{-f}, \tag{3}$$

$$S = 1 - e^{-s}. \tag{4}$$

Dividing both sides of the equation (1) with the total work force $W \equiv E_t + U_t$ yields

$$\frac{\dot{U}(t)}{W} = s \cdot \frac{W - U(t)}{W} - f \cdot \frac{U(t)}{W}, \tag{5}$$

or

$$\dot{u}(t) + (s + f) \cdot u(t) = s, \tag{6}$$

where $u(t) = U(t)/W$ is the rate of unemployment,

$e(t) = E(t)/W$ is the employment rate.

The equation (6) is a linear differential equation of order one with constant coefficients. The solution to this equation can be obtained by standard methods and takes the following form:

$$u(t) = \frac{s}{s + f} + c \cdot e^{-(s+f)\cdot(t-t_0)}.$$

For $t = t_0$, we have $c = u(t_0) - s/(s + f)$ and therefore:

$$u(t) = u(t_0) \cdot e^{-(s+f)\cdot(t-t_0)} + \frac{s}{s + f} \cdot \left(1 - e^{-(s+f)\cdot(t-t_0)}\right).$$

From this equation, it is easily obtained

$$u(t_0 + 1) = u(t_0) \cdot e^{-(s+f)} + \frac{s}{s + f} \cdot \left(1 - e^{-(s+f)}\right). \tag{7}$$

The equation (7) holds for any $t = t_0$, so we have a following difference equation of order one:

$$u_{t+1} = c + \rho \cdot u_t, \tag{8}$$

where $c = (s/(s + f)) \cdot \left(1 - e^{-(s+f)}\right)$,

$\rho = e^{-(s+f)}$.

Alternatively, we could express this difference equation in a familiar form:

$$(u_{t+1} - \bar{u}) = \rho \cdot (u_t - \bar{u}), \tag{9}$$

where $\bar{u} \equiv \frac{s}{s+f} = \frac{c}{1-\rho}$ is a stationary unemployment rate,

$(u_t - \bar{u})$ is a cyclical component of unemployment used to measure business cycles.

3 Data

The empirical counterpart to the variable u_t is a standardized unemployment rate in the Eurozone in fixed composition[3] with monthly frequency from 1999m1 to 2014m2. This time series is freely available at the European Central Bank (ECB) database:
http://sdw.ecb.europa.eu/quickview.do?SERIES_KEY=132.STS.M.I6.S.UNEH.RTT000.4.000

The data from each country were aggregated and then seasonally adjusted. These procedures were done by Eurostat which is the original source of the time series. The principles of aggregation and seasonal adjustment are described in „ESS Guidelines on Seasonal Adjustment“, which can be found at:
http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-09-006/EN/KS-RA-09-006-EN.PDF.

Data for this time series are available also before the year 1999, but these data are only estimates since the Eurozone originated in 1st January, 1999. For this reason, data before 1999 will not be used in the econometric analysis.

The question also arises whether or not to include data for periods before the economic crisis. The argument for not including these data is that economic crisis has dramatically changed the process generating the data. On the other hand, not including data before the economic crisis would mean the loss of many observations that could potentially contain valuable information. In the next chapter, all the data since 1999 will be used and the economic crisis will be modeled as a change of (selected) parameters.

Time series for the unemployment rate in the Eurozone is illustrated in the figure 1. Additionally, two important dates characterizing the breakpoint of the economic crisis are also illustrated and will be further discussed in the next chapter.

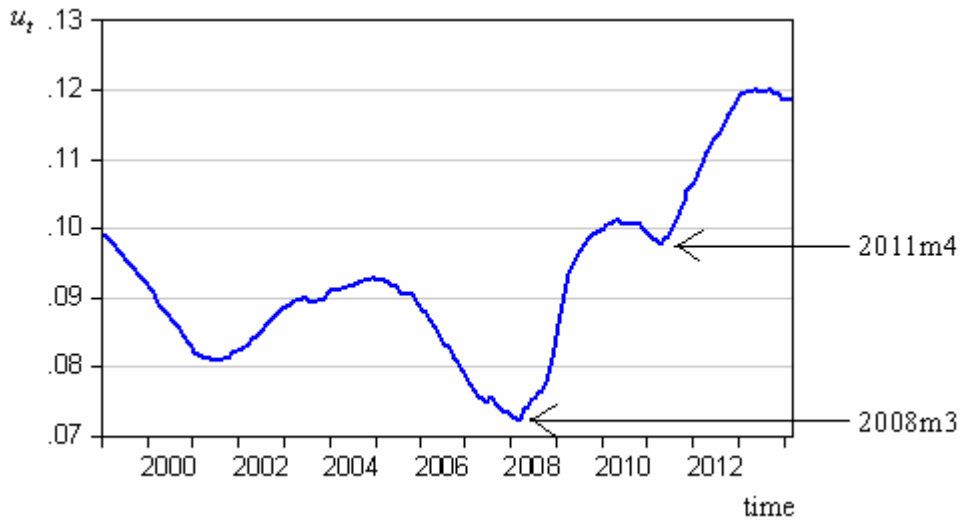


Figure 1 The unemployment rate in the Eurozone and the beginning of the economic crisis in 2008

4 Econometric model and its estimation[4]

The econometric estimation of the dynamic model of unemployment (8) derived in the 2. chapter requires random error to be added to the equation. Stochastic specification will be done in a standard following manner:

[3] Fixed composition means that the whole time series contains data from all the 17 countries of the Eurozone, despite the fact that not all states were members of the Eurozone from 1999.
[4] The econometric estimation of the model was performed in Eviews 8.

$$u_{t+1} = c + \rho \cdot u_t + \varepsilon_{t+1}, \tag{10}$$

where ε_{t+1} is a random error[5].

The rate of unemployment has been rising since the beginning of the economic crisis in 2008 which can be seen from the above figure 1. The rising trend was started in the second quarter of 2008 and further accelerated in the fourth quarter of 2008. There is also one additional important date 2011m4 illustrated in the figure1. At this month, there was a sudden change in the dynamics of u_t and unemployment rate began to rise again quite steeply.

The Chow breakpoint test for stability of the equation (10) with two proposed breakpoints 2008m3, 2011m4 was performed. The null hypothesis of the stability was easily rejected according to this test. The change of the reduced parameters c, ρ is motivated by the change of the structural[6] parameters s, f. Empirical evidence supporting the hypothesis of rising s and falling f after the economic crisis in the European countries can be found in Arpaia, Curci [1].

The effects of the economic crisis will be modeled as a change of the parameters. Not all of the parameters will be allowed to change however, since the model would then be overparameterized. To see this, recall that $\overline{u} = c/(1-\rho)$. The higher level of stationary unemployment rate $\overline{u}$ typical for the economic crisis[7] can thus be explained by the model (10) either by higher value of c or by higher value of ρ. For this reason, only one of these parameters (for example c) will be allowed to change:

$$u_t = c_t + \rho \cdot u_{t-1} + \varepsilon_t, \tag{11}$$

where $$c_t = \begin{cases} c^0, & \text{for } t = 1999m1,...,2008m3, \\ c^1, & \text{for } t = 2008m4,...,2011m4, \\ c^2, & \text{for } t = 2011m5,...,2014m2. \end{cases}$$

The regression model (11) was estimated by ordinary least squares (OLS). The serial autocorrelation of the residuals was however detected[8]. The partial autocorrelation function had statistically significant spikes at the order of one and two. The serial correlation of random errors of the order two will therefore be explicitely modeled:

$$\varepsilon_{t+1} = \lambda_0 \cdot \varepsilon_t + \lambda_1 \cdot \varepsilon_{t-1} + \varepsilon^*_{t+1}, \tag{12}$$

where ε^*_{t+1} is an i.i.d. random error.

The dynamic model of unemployment consisting of equations (11), (12) was estimated in Eviews by the following command:[9]

ls u bin_0 bin_1 bin_2 u(-1) ar(1) ar(2).

The correlogram of the residuals ε^*_{t+1} from this regression model was already in line with the hypothesis that ε^*_{t+1} is white noise. This result was further confirmed by the Breusch-Godfrey statistical LM test for the serial correlation of ε^*_{t+1}.

[5] It is not assumed that the random error ε_{t+1} is an i.i.d. sequence. These random errors will turn out to be autocorrelated with its past values. Proper form of the autocorrelation will be determined empirically.

[6] The parameters s, f aren't truly "structural" from this point of view, because structural parameters are normally considered to be constant over time.

[7] Towards which the unemployment rate u_t has been rising since 2008.

[8] Standard Durbin-Watson statistic wasn't used because of the lagged dependent variable among regresors.

[9] Eviews estimation technique for this command is the Marquardt nonlinear least squares algorithm. The "bin_i" variables are binary variables taking on values either zero or one.

Heteroskedasticity was also present according to the White test. Standard errors of the estimated parameters (and associated t-statistics) were therefore computed on the basis of heteroskedasticity consistent covariance (White) estimator.

The estimated parameters as well as associated t-statistics are summarized in the table 2:

coefficient	c^0	c^1	c^2	ρ	λ_0	λ_1
estimated value	0.0052	0.0066	0.0078	0.936	0.453	0.404
t-statistic	0.006	0.003	0.002	0.000	0.000	0.000

Table 1 Results of econometric estimation of the model (11), (12)

The overall fit of this regression model is extremely high ($R^2_{adj} = 0.9989$) and the standard error of the random variable ε^*_{t+1} is extremely low ($std\left(\varepsilon^*_{t+1}\right) = 0.00042$).

5 Forecasting

The estimated model (11), (12) will be used to forecast the unemployment rate in the Eurozone. Fitted values correspond to one-month-ahead in-sample forecasts. These forecasts are very precise and practically correspond to true data[10] as can be seen from $R^2_{adj} = 0.9989$. In practice, however, one-year-ahead forecasts are more relevant. These forecasts were generated dynamically[11] for the period of economic crisis and are illustrated together with the 95 % confidence band in the following figure 2:

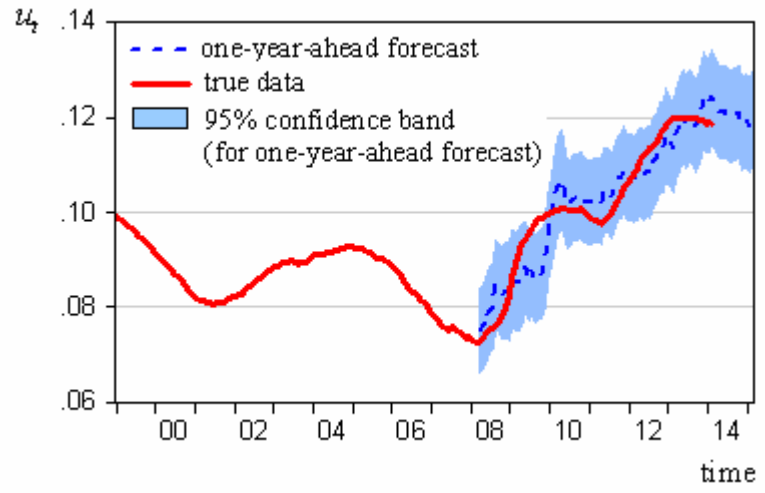


Figure 2 One-year-ahead forecasts of the unemployment rate u_t obtained from the model (11), (12)

The forecasts are compared to the true data for the unemployment in the Eurozone. Generated were in-sample forecasts (for periods from 2008m3 to 2014m2) as well as out-of-sample forecasts (for periods from 2014m3 to 2015m2). Dynamic one-year-ahead forecasts are much less precise. These results are nonetheless quiet common and can be considered to be a general rule rather than a shortcoming of the formulated model. More ARMA terms in the equation (12) for the random error as well as more lags of the variable u_t in the equation (12) did not improve the forecasting ability of the model.

The model predicts that the unemployment rate in the next year will become slightly lower. The model also states that the stationary rate of unemployment in the Eurozone for the final periods (2011m5,…,2014m2) is

[10] If these forecasts were drawn together with the true data, the two lines wouldn't be recognizable by eye.

[11] Dynamic forecast for future period $t+12$ made in time t uses (apart from the estimated parameters) only values of the variables, which are known at time t. It was however assumed that the values of the binary variables "bin_i" were known in advance.

$\hat{c}^2/(1-\hat{\rho}) = 0.0078/(1-0.936) = 0.12$. The long-run model prediction for the unemployment rate in the Eurozone is therefore equal to 0.12.

The word of caution should however be made. The model is not designed to forecast into a distant future. This particularly holds for the periods of current economic crisis. The possibility of another future breakpoint should especially be considered when applied this model in practice. There are economic reasons for another breakpoint to occur in the future. Lots of countries in the Eurozone (Italy, Spain, Greece) have currently serious problems with the public debt. These problems will require cuts in the future government expenditures, which will reduce aggregate demand for final products. This will lead to reduction of the aggregate demand for labor and thus also to rising unemployment.

6 Conclusion

Dynamic linear model of labor market was formulated and used to forecast unemployment rate within the Eurozone. One-month-ahead static forecasts practically coincide with the true data. Dynamic one-year-ahead forecasts are less precise, which is however a common feature and not a shortcoming of the formulated model. In-sample forecasting performance of the model can thus be considered to be performing reasonably well. One-year-ahead out-of-sample forecasts were also generated. The unemployment rate should become slightly lower during these periods according to these forecasts. Out-of-sample forecasts should nonetheless be interpreted with caution as the possibility of another future breakpoint during this period was not considered.

Various modifications within the modeling framework turned out not to improve the forecasting ability of the model. Better forecasting model would therefore have to include additional variables and (or) would have to be nonlinear. Nonlinearity as well as other variables could be introduced into the model by making probability of finding (loosing) a job endogenous. The method of measuring these probabilities is described in Shimer [6]. This methodology was applied for example by Elsby et al. [4], Barnichon [2]. Probability of finding a job could be made endogenous along the lines of the Pissarides [5] model. Dynamics of the probability of loosing a job is commonly assumed to be driven by exogenous random errors. These assumptions could form a basis for improvement of the formulated model.

Acknowledgements

Financial support of VŠE IGA F4/24/2014 is gratefully acknowledged by the author.

References

[1] Arpaia, A., Curci, N.: EU labour market behaviour during the great recession. *European Economy* no. 405, Brussels: European Commission, 2010.
[2] Barnichon, R.: Vacancy posting, job separation and unemployment fluctuations. *Journal of Economic Dynamics and Control*, vol. 36, issue 3, March 2012, 315-330.
[3] Barnichon, R., Nekarda, C. J.: The ins and outs of forecasting unemployment: using labor force flows to forecast the labor market. *Finance and Economics Discussion Series* of the Federal Reserve System no. 2013-09.
[4] Elsby, M. W. L., Hobijn, B., Sahin A. S.: Unemployment Dynamics in the OECD. Federal Reserve Bank of San Francisco Workong paper, 2011.
[5] Pissarides, C.: *Equilibrium Unemployment Theory*, 2nd edition, MIT Press, 2000.
[6] Shimer, R.: The cyclicality of hires, separations and job-to-job Transitions. *Federal Reserve Bank of St. Louis Review*,, July/August 2005, 87(4), 493-507.
[7] Shimer, R.: Reassessing the ins and outs of unemployment. Unpublished manuscript, University of Chicago, 2005.

Group possibilistic risk aversion in fuzzy pay-off method

Mikael Collan[1], Mario Fedrizzi[2], Pasi Luukka[3]

Abstract. Giga-investments are large industrial investments that have long construction times, long economic lives, and that are to a large degree irreversible. These characteristics make ex-ante giga-investment analysis difficult and require methods that can consider estimation of imprecision and the value of flexibility. Fuzzy pay-off method is a recently introduced profitability analysis tool, based on using managerial cash-flow scenarios estimated by a group of experts, to form a fuzzy (possibilistic) pay-off distribution for an investment that is compatible with the requirements of giga-investments. Due to their large size, most giga-investments require external financing that most often forces the companies, making the giga-investment, to acquire insurance coverage to bring down the idiosyncratic risk of the project to attract funding. Defining an insurance strategy requires an estimation of the risk as perceived by the group of experts (managers) who are supposed to be risk-averse. Accordingly, this paper analyzes the effect of the risk aversion in the possibilistic setting, relevant to giga-investments, in a multi-expert decision making context.

Keywords: pay-off method, possibilistic risk, group risk aversion and premium.

JEL Classification: G2, M2

1 Introduction

Giga-investments are very large industrial real investments that have a long construction time, a long economic life, and a high degree of irreversibility [14]. Giga-investments are "stationary" investments that tend to lock the strategy of the company undertaking them to the location, the technology, the quality, and to the capacity. The long construction time and the long economic life of the giga-investments make the accurate estimation of the future cost and revenue cash-flows difficult. The inaccuracy of estimation can come from a number of separate sources and combinations of sources, [38] divides the overall uncertainty facing managers to uncertainty from the general environment, the industry, and from organizational factors and [40] presents how different components of uncertainty affect capital budgeting practices. It is also rather clear that "as the time span increases, it is more likely that large changes will occur in the environment" [3] and all unexpected changes that happen during the economic life of giga-investments work to increase the uncertainty and the difficulty in, ex-ante, accurately perceiving the value of giga-investments.

Some giga-investments may have the ability to steer the markets by adjusting their production capacity; this ability gives managers a chance to try to actively optimize the value of the investment. Steering ability means that managers running giga-investments can make decisions based on private information that is not stochastic – steering ability is valuable and has an effect on the value of the giga-investment [30]. Any ability to steer the giga-investment, internal (such as changing the investment itself) or external (such as steering the markets) is made especially valuable by the high degree of irreversibility connected to giga-investments; "investments have a degree of irreversibility whenever they have attributes that make the capital specific to the firm, a product, or an industry, or else costly enough to move and relocate that the value of the capital becomes effectively tied to its original use. As a giga-investment cannot effectively be exited, it is a fair assumption that the companies and managers in charge of giga-investment decision-making dedicate a lot of attention to the risks involved in any giga-investment and the managerial flexibility that will allow them to steer these investments during their lifetime. Managerial flexibility of this kind is often referred to as real options [2]. Understanding the profitability of giga-investments correctly means that the analysis used must consider the value of managerial flexibility.

Possibility theory and fuzzy numbers are a natural fit in modeling giga-investment profitability, because they can be used to model imprecise investment cash-flows [35]. Possibility theory is also compatible with the fact that most of the information regarding giga-investment cash-flows commonly comes from managers and is based on normative expert assessment. In this paper we use the pay-off method [17] to model the profitability of giga-

[1] Lappeenranta University of Technology, School of Business, Skinnarilankatu 34, FI-53850 Lappeenranta,Finland, mikael.collan@lut.fi .
[2] University of Trento, Department of Industrial Engineering, Via Mesiano 77, I-38123 Trento, Italy, mario.fedrizzi@unitn.it.
[3] Lappeenranta University of Technology, School of Business, Skinnarilankatu 34, FI-53850 Lappeenranta,Finland, pasi.luukka@lut.fi.

investments, the method combines the use of managerial cash-flow scenarios with fuzzy logic and is able to consider real option value.

The majority of giga-investments are made with the support of external financing and financiers are understandably keen to ensure that any incidents that affect the productivity of a business, to which they have lent money, can recover as quickly as possible and with minimum impact on the revenue stream and the ability to repay the debt. Consequently, they will frequently insist, as a condition of financing that a company transfers as much risk as is commercially viable into the insurance markets. Financiers and their advisors will generally also take an interest in supplementary insurance coverage, whether compulsory or not, as a matter of wider professional due diligence. Minor project delays and contractual disputes can negatively impact on the delivery and management of an asset and it therefore makes sense to make sure that insurance coverage that can mitigate some of this risk is indeed purchased.

In this paper we examine how number of experts from the firm planning the giga-investment, as a group, measure (estimate) the project risk. We assume that individual risks are allocated in a Pareto-efficient way in the group. We suppose that the heterogeneity of individual beliefs does not come from asymmetric information, because the involved experts have access to the same information, but rather from intrinsic differences in how they view the project; mostly due to diverse levels of expertise and to varied knowledge base (e.g., technical / economic). The contribution of this paper is to illustrate how, in the presence of imprecise, vague, and even incomplete information, it is possible to create credible and relevant estimates for giga-investment risks and to diversify them (thus enabling the very existence of these large risky investments). From the possibilistic risk aversion theory the main contribution consists in introducing a totally new representation of group risk premium depending on individual possibilistic risk premiums previously introduced in [26].

2 The fuzzy pay-off method

Fuzzy pay-off method is a recently introduced method for investment profitability analysis that is based on using managerially generated cash-flow scenarios as a basis for creating a possibilistic (fuzzy) pay-off distribution for an investment [17], [15]. The method includes a procedure for the calculation of real option value, directly from the pay-off distribution by using the well-known option valuation logic that underlies all modern option valuation methods. The pay-off method has been previously used in the context of investment analysis for patents and IPR ([18] and [16] , mergers and acquisitions [19], research and development [32], [20], and giga-investments [14] .

Analysis with the fuzzy pay-off method starts by managerially generating cash-flow scenarios for an investment, usually three or four scenarios are used. When three scenarios are used, the scenarios correspond to the "best estimate" that is the most likely outcome of the giga-investment according to the managers, "maximum possible" and "minimum possible" scenarios. It is natural to expect that in the case of giga-investments the cash-flow scenarios are created by using information from multiple experts. The process can be of the type used in [16], where each expert creates three cash-flow scenarios, or the kind, where all experts together form three cash-flow scenarios. When giga-investments are discussed the second way is perhaps more relevant, because the group of experts involved in cash-flow estimation most likely represent different disciplines within the giga-investment whole, such as engineering, production management, financing, marketing, sales, and general management and not everyone of the experts may have a holistic view of the cash-flows. The giga-investment cash-flow scenarios are constructed from the knowledge of many experts, usually from within the company, the same approach is used, e.g, in [36].

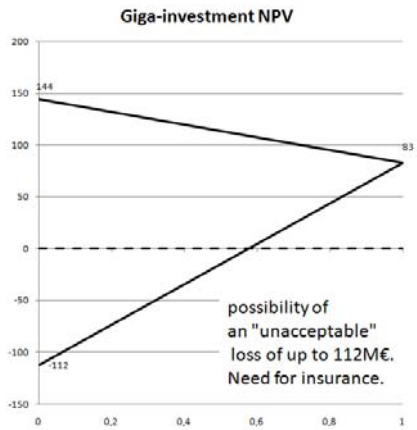


Figure 1 Pay-off distribution for a giga-investment

After the cash-flows are constructed the present value (PV) for each cash-flow and consecutively the net present value (NPV) for each scenario are calculated. When three scenarios are used the resulting scenario NPVs are used to form a simple pay-off distribution (see [17], [15]). It is the estimated possible downside risk, clearly visible in Figure 1 that is the primary cause for financiers to require the acquisition of risk cover from the markets to ensure the pay-back of their loans. The downside risk must be brought to an acceptable level and in practice it is seldom necessary to make the whole downside go away: by buying enough risk coverage from the markets (insurance) the downside can be capped to an acceptable level and the firm can consider the investment under risk neutrality by replacing the "catastrophic" cost of downside materializing by the cost of insurance. From the valuation point of view this means that the valuation of giga-investments can happen under risk neutrality after an insurance position is put into effect. The question that is raised is the correct or "optimal" level of insurance needed.

3 Possibilistic risk aversion and premium

In classical decision theory preferences are encoded by a utility function and alternatives are ranked according to expected utility criterion assuming that uncertainty is represented by a probability distribution that assigns probability values to each possible state [40]. When the available information is partial and/or based on subjective judgments of experts it is difficult to derive an appropriate probability distribution and therefore some non-probabilistic models of uncertainty such as Dempster-Shafer's theory, non-monotonic logic and possibility theory look more suitable as a basis for qualitative decision making.

Let us remark that the issue of how to choose the most suitable measure for representation and handling of uncertainty generated by subjective judgments expressed by experts has been harshly debated along almost the last thirty years and is still on the agenda of decision making processes' modelers. Since it's out of the scope of this paper to be focused on this kind of problem, the interested reader is directed to [7]. Here we assume that the possibilistic framework [41] is the most suitable and accordingly the available information is represented using fuzzy numbers and the choice of the best investment alternative(s) are based on the possibilistic counterpart of von Neumann and Morgenstern's expected utility criterion, as introduced by [22]. It is realistic to expect and assume that managers making a decision with regards to a large investment that by failing would potentially bankrupt the whole company are risk averse. Therefore, individual utility functions of these managers are strictly increasing and concave. Henceforth we identify an expert with her/his utility function, and the risky asset to evaluate is described by a triangular fuzzy number $\tilde{X}$. The problem of measuring individual risk aversion under a probabilistic framework was first addressed by [39] and [4], and then extended into several different decision making contexts (see, e.g., [23], [8],[21],[12] and [6]). Regarding giga-investments, let us remark that little attention has been paid to the problem of measuring risk aversion, except for, e.g., [5] and [13].

In [39] Pratt addressed the problem of measuring risk aversion considering a decision maker with assets x and utility function for money $u(x)$. he introduced the function $r(x)=-u''(x)/u'(x)$ as a measure of local risk aversion. Then he introduced the risk premium π assuming that the decision maker would be indifferent between receiving a risky asset Z and receiving the non-random amount $E(Z)-\pi$. The risk premium depends on x and on the distribution of Z, and accordingly will be denoted $\pi(x,Z)$. Then, by the expected utility criterion

$$u\left(x + E\left(z\right) - \pi\left(x, Z\right)\right) = E\left\{u\left(x + Z\right)\right\} \tag{1}$$

If $E\{u(x + Z)$ exists and is finite, then $\pi(x, Z)$ exists and is unique since $u(x + E(Z) - \pi(x, Z))$ is a strictly decreasing function of π. Then, he established a relation between π and $r(x)$ considering the risk premium for a small, actuarially neutral risk Z, i.e. a $\pi(x,Z)$ for $E(Z)=0$ and small variance σ_Z^2. Expanding $u(x)$ around x on both sides of (1) it comes out that the decision's maker risk premium is approximately $r(x)$ times half the variance of Z. We assume now that the ith expert's absolute risk aversion is represented by $r_i(x) = -\frac{u_i''}{u_i'}$ and is used to determine the individual possibilistic risk premiums as in [26], [27] and [28]. Accordingly, $r_i(x)$ measures the possibilistic absolute risk aversion of ith expert with respect to the uncertain payoff distribution as represented by $\tilde{X}$ and the corresponding risk premium. To determine the possibilistic risk premiums we will use the notions of possibilistic expected value and possibilistic variance as introduced by [9], [10], [25]. The possibilistic risk premium associated to ith expert, given $u_i(x)$ and the weighting function f, is defined by [26] as

$$u\left(E_f\left(\tilde{X}\right) - \pi_i\right) = E_f\left(u_i\left(\tilde{X}\right)\right) \tag{2}$$

By using second degree Taylor's formula, the ith expert's possibilistic risk premium π_i can be represented as

$$\pi_i \approx -\frac{1}{2}\sigma_f^*\left(\tilde{X}\right)\frac{u_i''}{u_i'} \tag{3}$$

where u_i'' and u_i' are calculated in $E_f(\tilde{X})$. It means, by analogy with the probabilistic case [39], that the measure of absolute risk aversion is here defined at the average level of payoff $E_f(\tilde{X})$. Assuming $\tilde{X}=(x,\alpha,\beta)$ and $f(t)=2t$, it's straightforward to show that [25]

$$E_f\left(\tilde{X}\right) = x + \frac{\beta - \alpha}{6} \tag{4}$$

and

$$\sigma_f^*\left(\tilde{X}\right) = \frac{(\alpha + \beta)^2}{18} - \frac{(\alpha - \beta)^2}{36} - \frac{\alpha\beta}{6} \tag{5}$$

Accordingly, the possibilistic risk premium of *i*th expert assumes the form

$$\pi_i = -\frac{1}{2}\left(\frac{(\alpha + \beta)^2}{18} - \frac{(\alpha - \beta)^2}{36} - \frac{\alpha\beta}{6}\right)\frac{u_i''\left(E_f\left(\tilde{X}\right)\right)}{u_i'\left(E_f\left(\tilde{X}\right)\right)} \tag{6}$$

4 Group possibilistic risk premium and insurance strategy

Suppose that a group of *n* experts (managers) faces the problem of valuating a giga-investment project using the fuzzy payoff method. The group members are uncertain about the payoff estimation, and here we assume that uncertainty is described in terms of a possibility distribution and that the group members agree on representing it as a triangular fuzzy number. The individual preferences on payoff domain's values are represented by utility functions for money $u_1(x), \ldots, u_n(x)$. We assume that each $u_i(x)$ belong to the class C^2, and is strictly increasing and concave, i.e. experts are risk averse. The first step of the group process consists in determining the group utility function $u_G(x)$. Expressing u_G as a function of the individual utility functions necessitates the consideration of lotteries involving $u_1(x), \ldots, u_n(x)$. In [31] Harsanyi proved that if certain conditions are satisfied, $u_G(x)$ can be expressed as a linear combination (with positive coefficients) of the individual utility functions. In [33] and [34] alternate sets of conditions were introduced, that lead to a group utility function that is a linear combination of $u_1(x), \ldots, u_n(x)$, as well as conditions that lead to more general forms for the group utility function. Here we assume that the linear aggregation rule introduced in [31] fits well the nature of the problem addressed and therefore

$$u_G(x) = \sum_{i=1}^{n} \lambda_i u_i(x) \tag{7}$$

with the additional conditions that the u_is are scaled from *0* to *1* and $\lambda_i>0$. The λ_i represents the weight granted to *i*th expert, e.g. decision power.

Henceforth we address the problem of how to determine the measure of risk aversion and the risk premium of the group of experts assuming that the group utility function is an affine aggregation of individual utilities functions, as in formula (7), and the assets are represented by a possibility distribution (triangular fuzzy number). Let us remark that the problem of risk aversion measure has been marginally addressed in probabilistic group decision making contexts and the main contributions are due to [24], [11] and [29]. Starting from formula (7) we determine the measure of group absolute risk aversion as

$$r_G(x) = \frac{u_G''(x)}{u_G'(x)} = -\frac{\sum_{i=1}^{n} \lambda_i u_i''(x)}{\sum_{i=1}^{n} \lambda_i u_i'(x)} \tag{8}$$

and from $u_i''(x) = -r_i(x)u_i'(x)$, $i=1\ldots n$, we obtain

$$r_G(x) = \frac{\sum_{i=1}^{n} \lambda_i r_i(x) u_i'(x)}{\sum_{i=1}^{n} \lambda_i u_i'(x)} \tag{9}$$

Since the duration of the decision process is quite short it's reasonable to assume constant individual absolute risk aversion, i.e. $r_i(x)=k_i$ with $k_i>0$, and therefore

$$r_G(x) = \frac{\sum_{i=1}^{n} \lambda_i k_i u_i'(x)}{\sum_{i=1}^{n} \lambda_i u_i'(x)} \tag{10}$$

A utility function of this kind is usually called CARA (Constant Absolute Risk Aversion) utility function. When ith expert's preferences exhibit CARA they can be represented by the utility function $u_i(x)=1-e^{-k_i x}$, where $k_i\,(>0)$ is the measure of absolute risk aversion and, accordingly, $u_G(x)=\sum_1^n \lambda_i(1-e^{-k_i x})$. Therefore, it's easy to prove that the group possibilistic risk premium takes the form

$$\pi_G = \frac{\sum_{i=1}^{n} \lambda_i \pi_i^2 e^{-k_i \tilde{u}}}{\sum_{i=1}^{n} \lambda_i \pi_i e^{-k_i \tilde{u}}} \tag{11}$$

where $\tilde{\mu} = E_f(\tilde{X})$. Putting $\lambda_i e^{-k_i \tilde{\mu}} = \alpha_i$ we obtain

$$\pi_G = \frac{\sum_{i=1}^{n} \alpha_i \pi_i^2}{\sum_{i=1}^{n} \alpha_i \pi_i} \tag{12}$$

Here the group possibilistic risk premium is represented as a linear homogeneous function of individual possibilistic risk premiums and depends as well on the way in which the individual decision powers influence the group utility functions.

5 Concluding remarks

A recent Accenture study [1] provides data that suggests companies are becoming more risk averse in their decisions related to developing very large investment projects i.e., giga-investments. This survey of 519 company executives in 12 industry sectors was conducted in France, the U.K., and the U.S. According to the study, 46% of the executives said their company had become more risk averse when considering new ideas. Accordingly, since due to their large size most giga-investments require external financing, companies are forced to acquire insurance coverage to bring down the idiosyncratic risk of the project to attract funding. Defining an insurance strategy requires an estimation of the risk as perceived by the group of experts (managers) who are in charge of the planning of the project. The measure of the effect of risk aversion on the insurance strategy depends on the risk premium the estimation of which is carried out starting from the individual risk premiums of the managers. In this paper, starting from the introduction of fuzzy pay-off method and of risk aversion measure and risk premium in a possibilistic framework we introduce a new extension to a group decision making settlement. Assuming that the group utility function is obtained as an affine aggregation of the individual utilities functions, we introduce the group risk aversion measure and the corresponding possibilistic risk premium.

References

[1] Accenture Study: *Innovation Efforts Falling Short Despite Increased Investment*. May 2013

[2] Amram, M. and Kulatilaka, N.: *Real Options: Managing Strategic Investment in and Uncertain World*. Harvard Business School Press, Boston, Mass., 1999

[3] Armstrong, S., and Crohman, M.: A comparative study for methods for long-term market forecasting. *Management Science* **19** (1972), 211-221.

[4] Arrow, K.J.: *Essays in the Theory of Risk Bearing*, North-Holland, Amsterdam, 1970

[5] Arrow, K.J.: Uncertainty and the evaluation of public investment decisions. *The American Economic Review* **60** (1970), 364-378.

[6] Bommier, A., Chassagnon, A., and Le Grand, F.: Comparative risk aversion: A formal approach with applications to saving behavior. *Journal of Economic Theory* **147** (2012), 1614-1641.

[7] Brunelli, M., and Fedrizzi, M.: (2010). The various facets of uncertainty: Remarks on the role played by probability and possibility. In: *Proceedings of ICAOR'10, 2nd International Conference on Applied Operational Research*, August 25-27, Turku, Finland.

[8] Caballè, J.: Mixed risk aversion. *Journal of Economic Theory* **71** (1996), 485-513.

[9] Carlsson, C., and Fuller, R.: On possibilistic mean value and variance of fuzzy numbers. *Fuzzy Sets and Systems* **122** (2001), 315–326.

[10] Carlsson, C., Fuller, R., and Majlender, P.: A possibilistic approach to selecting portfolios with highest utility score. *Fuzzy Sets and Systems* **131** (2002), 13–21.

[11] Chambers, C., and Echenique, F.: When does aggregation reduce risk aversion? *Games and Economic Behavior* **76** (2012), 582-595.
[12] Chen, A., Pelsser, A., and Vellekoop, M.: Modeling non-monotone risk aversion using SAHARA utility functions. *Journal of Economic Theory* **146** (2011), 2075-2092.
[13] Chronopoulos, M., De Reyck, B., and Siddiqui, A.: Optimal investment under operational flexibility, risk aversion, and uncertainty. *European Journal of Operational Research* **213** (2011), 221-237.
[14] Collan, M.: Valuation of industrial giga-investments: theory and practice. *Fuzzy Economic Review* **XVI** (2011), 21-37.
[15] Collan, M.: *The Pay-Off Method: Re-Inventing Investment Analysis*, CreateSpace Inc., Charleston, NC, USA, 2012
[16] Collan, M., M. Fedrizzi, and Luukka, P.: A multi-expert system for ranking patents: An approach based on fuzzy pay-off distributions and a TOPSIS–AHP framework. *Expert Systems with Applications* **40** (2013), 4749–4759.
[17] Collan, M., R, Fullér, and Mezei, J.: Fuzzy pay-off method for real option valuation. *Journal of Applied Mathematics and Decision Systems* **2009** (2009)
[18] Collan, M., and Heikkilä,M.: Enhancing patent valuation with the pay-off method. *Journal of Intellectual Property Rights* **16** (2011), 377-384.
[19] Collan, M., and Kinnunen, J.: A procedure for the rapid pre-acquisition screening of target companies using the pay-off method for real option valuation. *Journal of Real Options and Strategy* **4** (2011), 115-139.
[20] Collan, M., and Luukka, P.: Evaluating R&D projects as investments by using an overall ranking from four new fuzzy similarity measure based TOPSIS variants. *IEEE Trans. on Fuzzy Systems* **21** (2013), 1-11.
[21] Crainich, D., and Eeckhoudt, L.: On the intensity of downside risk aversion. *Journal of Risk and Uncertainty* **36** (2008), 267-276.
[22] Dubois, D., Godo, L., Prade, H., and Zapico, A.: Making decisions in a qualitative setting: From decision under uncertainty to case-based decision. In: *KR´98: Principles of Knowledge Representation and Reasoning* (Cohn, A.G., Schubert, L., and Shapiro, S.C. eds.). Morgan Kaufman Publishers, Inc, San Francisco, 1998, 594-605.
[23] Duncan, G. T.: A matrix measure of multivariate local risk aversion. *Econometrica*, **45** (1977), 895–902.
[24] Eliashberg, J., Winkler, R.: Risk sharing and group decision making. *Management Science* **27** (1981), 1221-1235.
[25] Fuller, R., and Majlender, P.: On weighted possibilistic mean and variance of fuzzy numbers. *Fuzzy Sets and Systems* **136** (2003), 365–374.
[26] Georgescu, I.: Possibilistic risk aversion. *Fuzzy Sets and Systems* **160** (2009), 2608–2619.
[27] Georgescu, I.: A possibilistic approach to risk aversion. *Soft Computing* **15** (2011), 795–801.
[28] Georgescu, I.: Computing the risk indicators in fuzzy systems. *Journal of Information Technology Research*, **5** (2012), 63-84.
[29] Gong, M., Baron, J., and Kunreuther, H.: Why do groups cooperate more than individuals to reduce risks? *Theory and Decision* **75** (2012), 101-116
[30] Harris, R.: On the choice of large projects. *The Canadian Journal of Economics* **11** (1978), 404-423.
[31] Harsanyi, J. C.: Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility, *J. Political Economy* **63** (1955), 309-321.
[32] Hassanzadeh, F., Collan, M., and Modarres, M.: A practical approach to R&D portfolio selection using fuzzy set theory. *IEEE Transactions on Fuzzy Systems* **20** (2012), 615-622.
[33] Keeney, R. L., and Kirwood, C. W.: Group Decision Making Using Cardinal Social Welfare Functions, *Management Science* **22** (1975), 430-437.
[34] Keeny, R., and Raiffa, H.: *Decisions with Multiple Objectives, Preferences, and Value Tradeoffs*. John Wiley, New York, 1976.
[35] Kuchta, D.: Fuzzy capital budgeting. *Fuzzy Sets and Systems* **111** (2000), 367-385.
[36] Mathews, S., Datar, V., et al.: A practical method for valuing real options: The Boeing approach. *Journal of Applied Corporate Finance* **19** (2007), 95-104.
[37] Miller, K.: A framework for integrated risk management in international business. *Journal of International Business Studies* **23** (1992), 311-331.
[38] Pratt, J.: Risk aversion in the small and in the large. *Econometrica* **32** (1964), 122–130.
[39] Verbeeten, F.: *The Impact of Uncertainty on Capital Budgeting Practices*. Willem-Jan van der Wolf, Nijmegen, The Netherlands, 2001.
[40] Von Neumann, J., and Morgenstern, O.: *Theory of Games and Economic Behaviour*. Princeton Univ. Press., Princeton, NJ., 1944.
[41] Zadeh, L.: Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems* **1** (1978), 3-28.

Are exchange rates in CEE countries driven by monetary fundamentals? Evidence from a panel approach

Marek A. Dąbrowski[1], Monika Papież[2], Sławomir Śmiech[3]

Abstract. The paper investigates whether nominal exchange rates in Central and Eastern European countries are driven by monetary fundamentals. The conventional monetary model is modified to include the difference in the relative price of non-tradables since this is a potentially important factor for exchange rate determination in emerging market economies. Using quarterly data spanning from 2001q1 until 2012q4 we detect cross-sectional dependence in our panel. Thus, in order to avoid erroneous inferences about cointegration and causality, we apply the methodology for non-stationary panels, which allows for cross-sectional dependence.
We find evidence of cointegration implied by the extended monetary model. Though estimated coefficients deviate slightly from those implied by the theory, their signs are correct. Granger causality tests reveal that: (i) in the long-term the exchange rates revert to the equilibrium relation identified; (ii) in the short-term exchange rate Granger-causes the differential in relative price of non-tradables and is Granger-caused by the relative money supply. The results are driven neither by episodes of high inflation nor by the global financial crisis.

Keywords: monetary exchange rate model, Central and Eastern European countries, cross-sectional dependence, panel cointegration, Granger causality.

JEL Classification: C33, E44; F41, F36
AMS Classification: 91B64, 91G70

1 Introduction

The exchange rate determination has become one of the major puzzles in international macroeconomics since Meese and Rogoff [12] demonstrated that a random walk model predicts exchange rates as well as any structural model in the short-run and Mussa [13] observed that the exchange rate volatility differed across exchange rate regimes even though volatility of macroeconomic fundamentals did not. Exchange rates were seen as instable and their fluctuations as disconnected from macroeconomic fundamentals.

Advancement in econometrics, however, has opened new avenues for research and equipped researchers with new econometric techniques. It was found that the null hypothesis of no cointegration between the nominal exchange rate and fundamentals was difficult to reject empirically because the available spans of data were relatively short and the standard tests had low power (see [21]). Thus, the natural way to overcome this problem was to use panel data[4]. The line of research based on panel cointegration methods turned out to be relatively successful in 're-connecting' an exchange rate with fundamentals both for advanced economies (see, e.g., [11], [5]) and emerging market economies (see, e.g., [6], [22], [23]).

Though studies on exchange rates in CEE countries that adopt this approach provide the evidence of cointegration between exchange rate and monetary fundamentals (e.g., [6], [22], [23]), their findings are rather tentative rather than conclusive. There are three problems that cast some doubt on the results. First, in these studies the first generation unit root tests (e.g. IPS or LLC) and the first generation panel cointegration tests (see [15]) are used. A common feature of such tests is the assumption of cross-sectional independence of panel units. If it is invalid the null hypotheses of non-stationarity and a lack of cointegration are rejected too often (see [2]). Second, the datasets used in the previous studies covered periods of high inflation. Since the monetary approach to the exchange rate 'performs fairly well when inflation is high' (see [10]), the previous results for CEE countries could be driven mainly by inflation differentials (see e.g. [22]). Third, though the monetary model is supposed to

[1] Cracow University of Economics, Faculty of Economics and International relations, Department of Macroeconomics, Rakowicka 27, 31-510 Cracow, Poland, e-mail: marek.dabrowski@uek.krakow.pl.
[2] Cracow University of Economics, Faculty of Management, Department of Statistics, Rakowicka 27, 31-510 Cracow, Poland, e-mail: papiezm@uek.krakow.pl.
[3] Cracow University of Economics, Faculty of Management, Department of Statistics, Rakowicka 27, 31-510 Cracow, Poland, e-mail: smiechs@uek.krakow.pl.
[4] An alternative was to use long spans of data (see [19]). This is not discussed due to data availability constraint for CEE countries.

explain exchange rate fluctuations in countries under relatively flexible exchange rate arrangements, panels used in the previous studies included countries with fixed exchange rate arrangements (e.g. Baltic states with currency boards were included in [22]).

In this paper we examine whether the exchange rates of CEE currencies (against the euro) are related to macroeconomic fundamentals as implied by the monetary model. Our contribution to the literature is threefold. First, we find evidence of long-run relationship between nominal exchange rate and fundamentals in CEE countries that is consistent with the monetary model and more reliable than the results of extant studies. Unit root and cointegration tests employed in this paper allow on cross-sectional dependence, the sample period is not contaminated with the episodes of high inflation and the panel consists of countries with relatively flexible exchange rate regimes. Second, the Granger-causality analysis reveals that the nominal exchange rate is related to fundamentals both in the long-run (it readjusts to the equilibrium relation after a shock), and in the short-run. Third, even though the global financial crisis (GFC) has had a relatively strong impact on CEE countries (in comparison with Asian and Latin American emerging market economies), our results are not driven by the recent crisis.

The paper is organised as follows. In the next section we sketch the monetary model of exchange rate determination. Section 3 briefly presents methodology applied in the analysis of cross-sectional dependence, cointegration and causality. Empirical results are discussed in Section 4. Conclusions are offered in the last section.

2 The monetary model of the exchange rate

The monetary model of the exchange rate determination has become something of a workhorse in the exchange rate literature (see [6]). It is used to explain movements in exchange rates both in advanced economies (see e.g. [12], [8]) and in emerging market economies (see e.g. [6], [22], [7]).

Its three building blocks include: money market equilibrium conditions (in the domestic and foreign economies), the absolute purchasing power parity and the uncovered interest rate parity (see [20]). Under expectations that are rational and not explosive ('no bubbles' condition), the solution for the log of nominal exchange rate $s_{i,t}$, defined as a price of domestic currency in terms of foreign currency, is

$$s_{i,t} = (1-b)\sum_{j=0}^{\infty} b^j \mathrm{E}_t v_{i,t+j} \tag{1}$$

where $i = 1,\ldots,N$ corresponds to cross-sectional units, $t = 1,\ldots,T$ denotes the time period, E_t is the expectation formed on the basis of an information set available in the time t, b is a function of an interest rate semi-elasticity of money demand, and $v_{i,t}$ stands for the monetary fundamentals:

$$v_{i,t} \equiv -\left(m_{i,t} - m_t^*\right) + \kappa\left(y_{i,t} - y_t^*\right) + \left(p_{i,t} - p_{i,t}^T\right) - \left(p_t^* - p_t^{T*}\right) \tag{2}$$

Thus, the exchange rate is determined by current and expected future levels of fundamentals: relative stock of money, $m_{i,t} - m_t^*$, relative output, $y_{i,t} - y_t^*$, and a difference in the relative price of non-tradables[5], $(p_{i,t} - p_t^*) - (p_{i,t}^T - p_t^{T*})$. All variables are in log form and the asterisk denotes a foreign variable.

Subtracting $v_{i,t}$ from both sides of (1), it is possible to demonstrate that (see e.g. [20]):

$$s_{i,t} - v_{i,t} = \sum_{j=1}^{\infty} b^j \mathrm{E}_t \Delta v_{i,t+j} \tag{3}$$

From (1) we know that if fundamentals are non-stationary $I(1)$ processes, the nominal exchange rate is non-stationary as well. Transformation given by (3) reveals that exchange rate and its fundamentals are cointegrated with cointegrating vector $(1,-1)'$. Thus, the implied long-run equilibrium relationship to be examined in the empirical section is:

$$s_{i,t} = \beta_1\left(m_{i,t} - m_t^*\right) + \beta_2\left(y_{i,t} - y_t^*\right) + \beta_3\left[\left(p_{i,t} - p_{i,t}^T\right) - \left(p_t^* - p_t^{T*}\right)\right] + \varepsilon_{i,t} \tag{4}$$

[5] Here we make use of the definition of the general price level as a weighted average of prices of tradables and non-tradables that implies: $p_t - p_t^T \equiv \alpha(p_t^N - p_t^T)$. The parameter κ is an income elasticity of money demand.

with $\beta_1 = -1, \beta_2 = \kappa > 0, \beta_3 = 1$ and the error term, ε_t.

3 Methodology and data

In this study we use a panel cointegration technique to test the long-term and short-term relationships between the nominal exchange rate and macroeconomic fundamentals. According to [1], the default assumption of independence between cross-sections seems to be inadequate both in the cointegration analysis and causality analysis. Incorrect cross-sectional independence assumptions may lead to erroneous causal inferences. Therefore, to test for the presence of such cross-sectional dependence in our data, we apply two different statistics, Lagrange multiplier (*LM*) (see [4]) and *CD* (see [17]), with the null hypothesis claiming no cross-sectional dependence.

The estimation of the panel cointegration requires testing whether the variables contain unit roots. We use second-generation panel CIPS unit root test proposed in [18]). To test for cointegration we used the second generation tests proposed in [24]. The equation for Westerlund cointegration tests can be formulated as follows:

$$s_{i,t} = \beta_0 + \beta_1 m^r_{i,t} + \beta_2 y^r_{i,t} + \beta_3 x^r_{i,t} + \varepsilon_{i,t} \tag{5}$$

where $i = 1,\ldots,N$ denotes the number of countries in the panel, and $t = 1,\ldots,T$ denotes the time period. The variables $s_{i,t}$, $m^r_{i,t} \equiv m_{i,t} - m^*_t$, $y^r_{i,t} \equiv y_{i,t} - y^*_t$, $x^r_{i,t} \equiv (p_{i,t} - p^*_t) - (p^T_{i,t} - p^{T*}_t)$, denote the natural logarithm of nominal exchange rate, the natural logarithm of money supply, the natural logarithm of real income, and the natural logarithm of the relative price of non-tradables for each country *i*, respectively. A superscript *r* means that the variable is a difference between the relevant home and foreign level. The monetary model of exchange rate determination implies that we should expect $\beta_1 < 0$, $\beta_2 > 0$ and $\beta_3 > 0$.

Given the presence of cointegration, we estimate the long-term parameters in the cointegrating vector. For this study, we use fully modified ordinary least squares FMOLS estimator (see [16]). The last step is to estimate a panel vector error correction model. This final representation is used to determine Granger causal relations between the variables. The two-step procedure of Engle-Granger (see [9]) is performed first by estimating the long-term model specified in Eq. (5) to obtain the estimated residuals, and then by defining the lagged residuals from Eq. (5) as the error correction term. The panel VECM can be written as follows:

$$\Delta s_{i,t} = \omega_{1i} + \sum_{j=1}^{p} \gamma_{11ij} \Delta s_{i,t-j} + \sum_{j=1}^{p} \gamma_{12ij} \Delta m^r_{i,t-j} + \sum_{j=1}^{p} \gamma_{13ij} \Delta y^r_{i,t-j} + \sum_{j=1}^{p} \gamma_{14ij} \Delta x^r_{i,t-j} + \theta_{1i} ECT_{i,t-1} + u_{1i,t}, \tag{6a}$$

$$\Delta m^r_{i,t} = \omega_{2i} + \sum_{j=1}^{p} \gamma_{21ij} \Delta s_{i,t-j} + \sum_{j=1}^{p} \gamma_{22ij} \Delta m^r_{i,t-j} + \sum_{j=1}^{p} \gamma_{23ij} \Delta y^r_{i,t-j} + \sum_{j=1}^{p} \gamma_{24ij} \Delta x^r_{i,t-j} + \theta_{2i} ECT_{i,t-1} + u_{2i,t}, \tag{6b}$$

$$\Delta y^r_{i,t} = \omega_{3i} + \sum_{j=1}^{p} \gamma_{31ij} \Delta s_{i,t-j} + \sum_{j=1}^{p} \gamma_{32ij} \Delta m^r_{i,t-j} + \sum_{j=1}^{p} \gamma_{33ij} \Delta y^r_{i,t-j} + \sum_{j=1}^{p} \gamma_{34ij} \Delta x^r_{i,t-j} + \theta_{3i} ECT_{i,t-1} + u_{3i,t}, \tag{6c}$$

$$\Delta x^r_{i,t} = \omega_{4i} + \sum_{j=1}^{p} \gamma_{41ij} \Delta s_{i,t-j} + \sum_{j=1}^{p} \gamma_{42ij} \Delta m^r_{i,t-j} + \sum_{j=1}^{p} \gamma_{43ij} \Delta y^r_{i,t-j} + \sum_{j=1}^{p} \gamma_{44ij} \Delta x^r_{i,t-j} + \theta_{4i} ECT_{i,t-1} + u_{4i,t}, \tag{6d}$$

where *p* is the optimal lag length(s) determined by the BIC, $ECT_{i,t}$ is the error correction term, that is the lagged residuals from the panel FMOLS estimation of the Eq. (5), and the error term $u_{i,t}$ is assumed to be i.i.d. with a zero mean and constant variance. The VECM is estimated using SUR techniques that allow for cross-sectional specific coefficient vectors and cross-sectional correlation in the residuals.

We use the balanced panel of quarterly data spanning from 2001:4 till 2012:4 for eight CEE countries: the Czech Republic, Hungary, Moldova, Poland, Romania, Serbia, Turkey and Ukraine. Nominal exchange rates are end-of-quarter observations from the IMF/CEIC database. They are expressed in euros per national currency, so an increase means an appreciation of national currency. Our measure of money stock is an aggregate M2 from the IMF/CEIC database. Real GDP is used to measure economic activity. The data come from the national sources for Moldova and Serbia, from the IMF/CEIC for Turkey and Ukraine, and from the Eurostat for other countries. The general price levels are measured with GDP deflators and are collected from the same sources as

GDPs. Producer price indices from the IMF/CEIC database are used as a proxy for the price of tradables. All variables are expressed as indices (2005 = 100) and specified in natural logarithms. The seasonal adjustment has been done by means of the Eviews default version of the Census X12 algorithm.

4 Empirical results

The analysis of panel data begins with testing for cross-sectional dependence. The results of two different cross-sectional dependence test statistics, LM (see [4]), and CD (see [17]), reject the null hypothesis of no cross-sectional dependence with *p* values lower than 1 percent for all variables included in our framework. Since we find evidence of cross-sectional dependence among the variables, we use the second-generation panel unit root tests (see [18]) and we apply the cointegration test developed in [24]. CIPS unit root test results for both 'constant' and 'constant and trend' specifications do not reject the null of unit roots for the panel at level for most variables, on the contrary, the differenced series are stationary. After establishing the nonstationarity of the time series, we test for the existence of the long-term relationship between the nominal exchange rate and macroeconomic fundamentals using the Westerlund cointegration test. G_a and G_t group test statistics show that the null hypothesis of no cointegration for at least one of the cross-sectional units is rejected in favour of cointegration at the 5 percent level. Similarly, P_a and P_t panel test statistics using the pooled information over all cross-sectional units indicate that the null hypothesis of no cointegration for the whole panel is rejected at the 10 percent level[6]. Overall, the results of cointegration tests reveal the existence of the long-term relationship between the nominal exchange rates and fundamentals in CEE countries even if the cross-sectional dependence is allowed for.

Since evidence confirms the long-term relationship, we next estimate long-term coefficients using the panel FMOLS. The FMOLS estimates of equation (5) are as follows:

$$s_{i,t} = 4.586 - \underset{(-4.717)***}{0.300}\, m^r_{i,t} + \underset{(2.649)***}{0.502}\, y^r_{i,t} + \underset{(6.776)***}{1.349}\, x^r_{i,t} \tag{7}$$

where the numbers in parentheses are *t*-statistics and variables.

Since variables are in natural logarithms, the estimated coefficients can be interpreted as long-run elasticities. All of them are highly significant and correctly signed. One per cent increase in the money supply triggers depreciation against the euro by 0.3 per cent in the long-run. Though it is smaller than the one per cent depreciation implied by the model, it is in line with estimates found in the literature. Beckmann et al. [3] estimate that for 18 OECD countries it is even lower (0.20), whereas the estimates for CEE countries provided in [6] range from 0.300 to 0.975 depending on the method applied.

One per cent rise in the relative output brings about an appreciation by 0.5 per cent. The finding that the income elasticity is greater than the money supply elasticity is in accordance with the results available for advanced economies in [3] or [11], but not with those for CEE countries obtained in [6] and [23]. The difference can be an outcome of inflation differentials dominance over the dataset, i.e. the flaw discussed in Section 1. Since we avoid this flaw, the estimate of β_2 is closer to the one for advanced economies, although still well below it (0.502 vs. 1.6 obtained in [3]).

One per cent increase of a differential in relative price of non-tradables results in a nominal appreciation against the euro of 1.3 per cent. This can be interpreted as a manifestation of the BS effect: a certain fraction of real currency appreciation takes a form of a nominal appreciation.

In the final step we estimate a panel vector error correction model in order to uncover the Granger causal relations between the variables. The Granger causality test is based on the model with a dynamic error correction term. Table 1 presents F-statistics and t-statistics which are useful in the short- and long-run Granger causality analyses respectively. Table 1 shows that the exchange rate is Granger caused by the relative money in the short run. On the one hand, it is a bit surprising since as pointed by Engel et al. [8] 'the models have little power to forecast exchange rate changes'. On the other hand, evidence of such a causality is also found in [8] for OECD countries. We conjecture that the exchange rate may possibly be considered an implicit target of monetary policy.

It is interesting to observe the causality running from exchange rate to the difference in relative price of non-tradables. As explained in [8] it is the causality of that type that is implied by the monetary model: an exchange rate is a reflection of expected future levels of fundamentals. Thus, it may be useful in predicting the fundamentals (assuming that the unobservable fundamentals are not a primary driver of an exchange rate).

[6] Detailed test results on cross-sectional dependence, unit roots and cointegration are available upon request.

Dependent variable	Source of causation (independent variables)				
	Short term - F-statistics				Long term - t-statistics
	Δs	Δm^r	Δy^r	Δx^r	ECT_{t-1}
Δs	-	9.675***	0.028	0.003	-5.113***
Δm^r	0.475	-	1.664	0.760	3.296***
Δy^r	1.741	6.122**	-	1.512	1.343
Δx^r	8.731***	0.005	0.529	-	3.552***

Notes: The null hypothesis is that there is no causal relationship between variables. Δ is the first difference operator. ECT_{t-1} represents the error correction term lagged one period. ***, **, * indicate statistical significance at 1, 5 and 10 percent level of significance, respectively.

Table 1 Panel Granger causality test results

The coefficient for the ECT in the exchange rate equation is significant, so the nominal exchange rate is related with fundamentals in the long-run as well. It indeed adjusts to the equilibrium relation implied by the model. Interestingly, the relative output seems to be weakly exogenous in the long-run (the coefficient on ECT is insignificant) and thus the return to the equilibrium relation after a shock involves changes in variables other than the relative output.

CEE countries as a group were severely hit by the crisis especially in comparison with Asian and Latin American emerging market economies. Thus, we introduce a crisis dummy in order to check whether the results of causality analysis are not an offshoot of the crisis. Results from Table 2 can be summarized as follows: even with the crisis dummy both short- and long-run relations remain unchanged. The exchange rate adjusts to the long-run equilibrium relation and in the short-run it is a Granger cause of the differential in relative price of non-tradables.

Dependent variable	Source of causation (independent variables)				
	Short term - F-statistics				Long term - t-statistics
	Δs	Δm^r	Δy^r	Δx^r	ECT_{t-1}
Δs	-	7.466***	0.192	0.163	-5.233***
Δm^r	0.381	-	1.493	0.632	3.329***
Δy^r	1.586	5.541**	-	0.640	1.006
Δx^r	7.254***	0.159	0.120	-	3.288***

Notes: Crisis dummy is equal to 1 for 2008:3-2009:4 and 0 otherwise. The null hypothesis is that there is no causal relationship between variables. Δ is the first difference operator. ECT_{t-1} represents the error correction term lagged one period. ***, **, * indicate statistical significance at 1, 5 and 10 percent level of significance, respectively.

Table 2 Panel Granger causality test results for a specification with a crisis dummy.

5 Conclusion

Using the second generation panel unit root and cointegration tests we have found that exchange rates and fundamentals in CEE countries are cointegrated as implied by the monetary model of exchange rate determination. This finding holds even though evidence of cross-sectional dependence among CEE countries is strong, the sample does not include periods of high inflation and the panel consists of countries with relatively flexible exchange rate regimes. Estimated elasticities deviate slightly from those implied by the theory but they are significant and correctly signed.

Granger causality tests reveal that in the long-term the exchange rates revert to the equilibrium relation identified. In the short-term the exchange rate is Granger-caused by the relative money supply and Granger-causes the differential in relative price of non-tradables. The results are not driven either by high inflation episodes or by the global financial crisis.

Acknowledgements

Supported by the grant No. 2013/11/B/HS4/00489 of the Polish National Science Centre.

References

[1] Bai, J. and Kao, C.: On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence. In: *Contributions to Economic Analysis* (Baltagi, B., ed.). Elsevier, 2006, 3-30.
[2] Banerjee, A., Marcellino, M., and Osbat, C.: Testing for PPP: Should we use panel methods? *Empirical Economics* **30** (2005), 77-91.
[3] Beckmann, J., Belke, A., and Dobnik, F.: Cross-section dependence and the monetary exchange rate model – A panel analysis. *North American Journal of Economics and Finance* **23** (2012), 38-53.
[4] Breusch, T., and Pagan, A.:The Lagrange multiplier test and its application to model specifications in econometrics. *Reviews of Economics Studies* **47** (1980), 239–253.
[5] Cerra, V., and Saxena, S.C.: The Monetary Model Strikes Back: Evidence from the World. *Journal of International Economics* **81** (2010), 184-196.
[6] Crespo-Cuaresma, J., Fidrmuc, J., MacDonald, R.: The monetary approach to exchange rates in the CEECs. *Economics of Transition* **13** (2005), 395-416.
[7] Dąbrowski, M. A., Papież, M., and Śmiech, S.: Monetary exchange rate model for the Central European countries: evidence from a panel approach. In: *Proceedings of 7th International Days of Statistics and Economics* (Loster, T., and Pavelka, T., eds.). Melandrium, Prague, 2013, 289-298.
[8] Engel, C., Mark, N. C., and West, K. D.: Exchange Rate Models Are Not as Bad as You Think. In: *NBER Macroeconomic Annual 2007*, (Acemoglu, D., Rogoff, K., and Woodford, M., eds.). Chicago University Press, Chicago, 2008.
[9] Engle, R.F., and Granger, C.W.J.: Cointegration and error correction: representation, estimation, and testing. *Econometrica* **55** (1987), 251–276.
[10] Flood, R. P., and Rose, A. K.: Understanding exchange rate volatility without the contrivance of macroeconomics. *Economic Journal* **109** (1999), 660-672.
[11] Groen, J.J.J.: Cointegration and the Monetary Exchange Rate Model Revisited. *Oxford Bulletin of Economics and Statistics* **64** (2002), 361-380.
[12] Mark, N.C., and Sul, D.: Nominal exchange rates and monetary fundamentals: Evidence form a small post-Bretton Woods panel. *Journal of International Economics* **53** (2001), 29-52.
[13] Meese, R., and Rogoff, K.: Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics* **14** (1983), 3-24.
[14] Mussa, M.: Nominal Exchange Rate Regimes and the Behavior of Real Exchange Rates: Evidence and Implications. *Carnegie-Rochester Conference Series on Public Policy* **25** (1986), 117–214.
[15] Pedroni, P.: Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics* **61** (1999), 653–670.
[16] Pedroni, P.:Fully modified ols for heterogeneous cointegrated panels. *Advances in Econometrics* **15** (2000), 93-130.
[17] Pesaran, M.H.: General diagnostic tests for cross section dependence in panels. *Cambridge Working Papers in Economics* No. 0435. Faculty of Economics, University of Cambridge, (2004).
[18] Pesaran, H.M.: A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics* **22** (2007), 265-312.
[19] Rapach, D. E., and Wohar, M. E.: Testing the monetary model of exchange rate determination: new evidence from a century of data. *Journal of International Economics* **58** (2002), 359-385.
[20] Sarno, L., and Taylor, M. P.: *The economics of exchange rates*. Cambridge University Press. Cambridge, UK, 2002
[21] Shiller, R. J., and Perron, P.: Testing the random walk hypothesis: power versus frequency of observation. *Economic Letters* **18** (1985), 381-386.
[22] Uz, I., and Ketenci, N.: Panel analysis of the monetary approach to exchange rates: Evidence from ten new EU members and Turkey. *Emerging Markets Review* **9** (2008), 57-69.
[23] Uz, I., and Ketenci, N.: Exchange rate determination: monetary approach in the new EU members and Turkey. *Applied Economics Letters* **17** (2010), 963-967.
[24] Westerlund, J.: Testing for Error Correction in Panel Data. *Oxford Bulletin of Economics and Statistics* **69** (2007), 709-748.

Factors of dissatisfaction with chosen study programme

Zuzana Dlouhá[1], Martin Dlouhý[2]

Abstract. This paper investigates determinants of dissatisfaction of the university graduates from the University of Economics, Prague, with chosen study programme. The data were obtained from the cross-sectional REFLEX survey that collected information regarding the labour market status of the graduates within the period 2008– 2012. We use the method of discrete binary choice model with explanatory variable describing the view of graduates if they would choose again the same study programme at the same institute of higher education. The data show that 35% of graduates are not satisfied with their study programme within five years after graduation. The findings show that main determinants of dissatisfaction with the study programme are gender, age, demanding study programme, employer familiar with study content, freedom to compose own programme, vocationally orientated, and horizontally mismatched at first job.

Keywords: discrete binary choice model, university education, REFLEX survey

JEL Classification: I23, J24
AMS Classification: 62P20

1 Introduction

Life is made up of everyday choices, some of them are easy to make, while others are more difficult and if some choices go wrong then spell regret. According to the analysis done by [5], the top six biggest regrets in life are (in descending order) education, career, romance, parenting, the self and leisure. First two of regrets – education and career are strongly dependent on the study programme choice, therefore it is important to understand what factors cause dissatisfaction (regret) with it [2]. Identifying the determinants of dissatisfaction with chosen study program can help universities better understand their weaknesses and strengths and identify areas for their improvement.

In this paper, we partially followed the approach introduced by [2] where determinants of regretting study programme by the graduates five years after their graduation in Spain and Netherlands were investigated such as education track and education-labour mismatch of tertiary education. The study [3] also confirms the role of over-education as a highly significant aspect affecting the field of study. On the other hand, there is also another approach introduced by [1] that education-labour mismatch is not the cause but the result of field regret and in the results from the Netherlands is confirmed that it occurred more often when skills were easily transferable across occupations.

The paper is structured as follows: the second chapter is dedicated to the used methodology and data description, in the third chapter we discuss the empirical results. In the fourth chapter we conclude our findings.

2 Methodology and data

We briefly introduce methodology used for determining the factors of dissatisfaction/regret with the graduated study programme within the following sub-chapter. Next we describe the analysed dataset and provide basic descriptive statistics of the variables incorporated into the model.

2.1 Methodology of binary choice model

We consider a class of binary response models of the form [6]

$$\mathrm{P}(y = 1|\boldsymbol{x}) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) = G(\beta_0 + \boldsymbol{x\beta}), \tag{1}$$

where G is a function taking on values strictly between 0 and 1: $0 < G(z) < 1$, for all real numbers z. For the estimation of response probabilities we use logit model, hence G is the logistic function:

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, e-mail: figlova@vse.cz.
[2] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, e-mail: dlouhy@vse.cz.

$$G(z) = \frac{e^z}{1+e^z} = \Lambda(z), \tag{2}$$

which has values between 0 and 1 for all real numbers z and this is the cumulative distribution function for a standard logistic random variable.
The expression P/(1 – P) presents *odds ratio* in favour of the occurrence of an event. If we take the natural log of this expression, the result is

$$L = ln\frac{\mathrm{P}}{1-\mathrm{P}} = \beta_0 + \boldsymbol{x\beta}, \tag{3}$$

and L (logit, hence the name logit model) is the natural log of the odds ratio linear in $\boldsymbol{x}$ and in parameters.

We use the maximum likelihood method for the estimation of the logit model. This method is consistent, normally distributed and efficient in large samples.

Marginal (partial) effects for continuous explanatory variable on $p(x) = \mathrm{P}(y = 1|\boldsymbol{x})$ is obtained from the partial derivative:

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{x\beta})\beta_j, \tag{4}$$

where $g(z) = \frac{\mathrm{d}G}{\mathrm{d}z}(z)$ is probability density function associated with G. Due to non-negativity of the density function, the partial effect of the x_j will always have the same sign as β_j.

Marginal (partial) effects for discrete explanatory variable on the probability x_k going from c_k to $c_k + 1$ is

$$G[\beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_k(c_k + 1)] - G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k c_k). \tag{5}$$

Standard errors of marginal effects (4) and (5) can be calculated using e.g. delta method [6].

To analyse the correctness of the model several tests will be run. For testing multiple restrictions the likelihood ratio (LR) as a twice the difference in the log-likelihoods is used:

$$LR = 2(L_{ur} - L_r), \tag{6}$$

where L_{ur} is the log-likelihood for the unrestricted model and L_r is the log-likelihood for the restricted model. As a goodness of fit we calculate the following McFadden's R^2:

$$\text{McFadden's } R^2 = 1 - (L_{ur} / L_r) \tag{7}$$

Hosmer-Lemeshow and Andrews χ^2 tests as another comparison of the fitted expected values to the actual values were used. To analyse the predictive accuracy of the model we evaluate the percentage of correct predictions with 0.5 as a cut-off value.

2.2 Data

The REFLEX survey (Research into Employment and professional FLEXibility) was held by the Education Policy Center, Charles University Prague, in 2013. This project is a large scale international project that has been carried out in 15 European countries and Japan. Altogether 21 public, and 15 private colleges and universities were participated on this survey in the Czech Republic. Sub-sample of graduates from the University of Economics, Prague consists of 1,704 respondents that graduated during years 2008–2012 and that are currently on the job market. Data was weighted by the proportion of faculties, type of study, gender, economical status and year of graduating. Due to lack of answers in some explanatory variables we use final dataset 1,482 graduates.

We constructed dummy explanation variable named *Dissatisfaction with chosen study programme* using results from the question *"Looking back, if you were free to choose again, would you choose the same study program at the same institute of higher education?"* We code answers with value 1 for the respondents that would study a different study programme in the same or a different institute, 0 for respondents satisfied with study programme. As explanatory variables we use gender (female = 1), age, six additional programme variables as answers to the question *"To what extent did the following descriptions apply to your study program?"*, with the scale 1 – not at all, to 5 – to a very high extent. Labour mismatching is presented by dummy variable named *Horizontally mismatched* in the first job where a graduate is considered as mismatched when his/her job is not related to the field of study of the programme [4]. Respondents were asked to exclude jobs they left within 4 months after graduation.

In Table 1 we present main descriptive statistics of all variables. We can observe that 34.68% of the respondents regret their study. From the program characteristics the highest rating (rated 3.6282) received *Easy program*

followed by *Academically prestigious.* More than 26% considered their field of study as not appropriate for their first job (variable *Horizontally mismatched*).

Variable	Mean	Median	Minimum	Maximum	Std. dev.
Dissatisfaction with chosen study programme	0.3468	0	0	1	0.4761
Female	0.6140	1	0	1	0.4870
Age	28.0459	28	21	53	3.5300
Programme characteristics					
Demanding[a]	2.4170	2	1	5	0.9283
Employer familiar with content[a]	2.7719	3	1	5	1.0455
Freedom to compose own programme[a]	2.6404	3	1	5	1.0787
Broad focus[a]	2.4636	2	1	5	0.9306
Vocationally orientated[a]	2.6849	3	1	5	1.0043
Academically prestigious[a]	2.7888	3	1	5	1.0266
Easy programme[a]	3.6282	4	1	5	0.9618
Horizontally mismatched	0.2638	0	0	1	0.4409

[a]Scale 1 to 5, age is continuous variable, the rest are all dummy variables.

Table 1 Descriptive statistics

3 Results

The main results of the binary logit estimation are presented below in Table 2. For all the estimations, econometric software EViews 8 was used. As odds ratios and marginal effects with standard errors are not included in EViews, we programmed own procedures. While the value of the coefficient does not give a clear interpretation of the relationship when using a binary logit model, the sign of the coefficient does have the same interpretation as with an ordinary least square model. From the results we observe that *Females* (significant at 5% level) have positive effect on the probability of dissatisfaction with chosen study programme, on the other hand *Age* of respondent lowers this probability. From *Programme characteristics* - except *Broad focus*, *Academically prestigious* and *Easy programme*, all the remaining variables are statistically significant at least at 10% level and they lead to more regret. Dummy variable *Horizontally mismatched* in the first job in general increases the likelihood of study programme regret.

Females are 1.3067 times more likely than males to regret their choice of study programme controlling all other variables on average when interpreting the odds ratio (3). The explanatory variables are also jointly highly significant (LR statistics = 202.5391, $p < 0.001$) despite the relatively low value of McFadden's pseudo R^2 of 0.1059.

Variable	Coefficient	Std. Error	Odds
Constant term	-2.3887	0.6573***	0.0917
Female (d)	0.2675	0.1253**	1.3067
Age	-0.0303	0.0181*	0.9702
Programme characteristics			
Demanding	0.1452	0.0847*	1.1563
Employer familiar with content	0.1063	0.0619*	1.1122
Freedom to compose own programme	0.1110	0.0617*	1.1174
Broad focus	-0.0318	0.0704	0.9687
Vocationally orientated	0.6597	0.0661***	1.9342
Academically prestigious	0.0050	0.0672	1.0050
Easy programme	-0.1001	0.0753	0.9047
Horizontally mismatched (d)	0.2488	0.1326*	1.2825

McFadden R-squared	0.1059	Mean dependent var	0.3468
S.D. dependent var	0.4761	S.E. of regression	0.4455
Akaike info criterion	1.1691	Sum squared resid	292.0003
Schwarz criterion	1.2084	Log likelihood	-855.3041
Hannan-Quinn criter.	1.1838	Deviance	1710.6080
Restr. Deviance	1913.1470	Restr. Log likelihood	-956.5736
LR statistic	202.5391	Avg. log likelihood	-0.5771
Prob(LR statistic)	0.0000	Mean dependent var	0.3468
Obs with Dep=0	968	Total obs	1482
Obs with Dep=1	514		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2 Determinants of dissatisfaction with chosen study programme

Table 3 reports the marginal effects and standard errors where all explanatory variables are set equal to their means in the sample. We can observe that marginal effects copies the sign of estimated regression coefficients from Table 1. Marginal effect for continuous variable *Age* measure the instantaneous rate of change – increase age by one year will produce 0.0067 decrease in the probability of dissatisfaction with chosen study programme as this variable is measured in small units (4). Marginal effect for dummy explanatory variable *Female* (equals to 0.0588) measures difference in probability of dissatisfaction with chosen study programme between females and males, holding all other variables at their means (5).

Variable	Marginal effect	Std. Error
Female (d)	0.0588	0.0275**
Age	-0.0067	0.0040*
Programme characteristics		
Demanding	0.0319	0.0186*
Employer familiar with content	0.0234	0.0136*
Freedom to compose own programme	0.0244	0.0136*
Broad focus	-0.0070	0.0155
Vocationally orientated	0.1450	0.0144***
Academically prestigious	0.0011	0.0147
Easy program	-0.0220	0.0165
Horizontally mismatched (d)	0.0547	0.0291*

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3 Marginal effects and of logit model

The results of Hosmer-Lemeshow and Andrews χ^2 tests in Table 4 indicate small difference between the fitted expected values and actual values. Therefore we cannot reject the model as providing an insufficient fit to the data.

Statistics	Value	Prob. χ^2
Hosmer-Lemeshow	9.6266	0.2922 (d.f. = 8)
Andrews	10.5156	0.3965(d.f. = 10)

Table 4 Goodness of fit evaluation

The results of the expectation-prediction evaluation test (or classification table) for estimated model are presented in Table 5. Result show that the estimated equation yields prediction-expectation values (or so called sensitivity) that are 35.99% and 43.27%, respectively, correct for the dissatisfaction with chosen study programme at a success cut-off rate 50% or higher. Overall, the estimated model predicts 69.50% of the observations with the value of specificity equals to 87.29%. The measure of predictive ability is reported in the row of Total Gain. The estimated equation is 4.18% better at predicting responses than the constant probability model. This change represents a 12.06% improvement over the 65.32% correct prediction of the default model. Analogous prediction results based upon expected value calculations are in the bottom part of Table 5.

Success cut-off: C = 0.5

	Estimated Equation			Constant probability		
	Dep = 0	Dep = 1	Total	Dep = 0	Dep = 1	Total
P(Dep = 1) <= C	845	329	1174	968	514	1482
P(Dep = 1) > C	123	185	308	0	0	0
Total	968	514	1482	968	514	1482
Correct	845	185	1030	968	0	1482
% Correct	87.29	35.99	69.50	100.00	0.00	65.32
% Incorrect	12.71	64.01	30.50	0.00	100.00	34.68
Total Gain*	-12.71	35.99	4.18			
Percent Gain**	NA	35.99	12.06			
E(# of Dep = 0)	676.43	291.57	968.00	632.27	335.73	968.00
E(# of Dep = 1)	291.57	222.43	514.00	335.73	178.27	514.00
Total	968.00	514.00	1482.00	968.00	514.00	1482.00
Correct	676.43	222.43	898.86	632.27	178.27	810.54
% Correct	69.88	43.27	60.65	65.32	34.68	54.69
% Incorrect	30.12	56.73	39.35	34.68	65.32	45.31
Total Gain*	4.56	8.59	5.96			
Percent Gain**	13.15	13.15	13.15			

* = change in "% correct" from default (constant probability) specification. ** = percent of incorrect (default) predictions corrected by equation.

Table 5 Expectation-prediction evaluation output of the model

4 Conclusion

The data show that almost 35% of graduates from the University of Economics, Prague, are not satisfied with their choice of study programme within five years after graduation. This study identified main determinants of dissatisfaction with the study programme that are: gender and age of the respondent and selected programme characteristics, such as demanding programme, employer familiar with study content, freedom to compose own programme, vocationally orientated programme, and horizontally mismatched at first job. The results of our study are similar to results found by Kucel and Vilalta-Bufí [2] in the term of the same sign and significance of estimated regression coefficients for the following variables: age of the respondent, vocationally orientated programme, and horizontally mismatched at first job.

In future, we plan to widen this study in terms of comparison of dissatisfaction with the choice of study programme with other universities in the Czech Republic. The comparison with other countries participating in the REFLEX survey is also possible. One can also compare results with previous years in order to evaluate possible developments in time.

Acknowledgements

The research is supported by the Internal Grant Agency of the University of Economics, Prague, project no. F4/19/2013.

References

[1] Borghans, L., and Golsteyn, B.: Skill transferability, regret and mobility. *Applied Economics* **39** (13), 1663-1677 (2007).

[2] Kucel, A., and Vilalta-Bufí, M.: Why do university graduates regret their study program? A comparison between Spain and Netherlands. *Working papers in Economics, Universitat de Barcelona. Espai de Recerca en Economia*, 279 (2012).

[3] Mora, T.: Why do higher graduates regret their field of studies? Some evidence from Catalonia, Spain. *Education Economics, Taylor & Francis Journals* **18** (2010), 93-109.

[4] Robst, J.: Education and job match: The relatedness of college major and work. *Economics of Education review* **26** (2007), 397-407.

[5] Roese, N. J., and Summerville, A.: What We Regret Most… and Why. *Personality and Social Psychology Bulletin* 31 (2005), 1273-1285.

[6] Wooldridge, J. M.: *Introductory Econometrics: A Modern Approach.* 5th edition, Cengage Learning, Mason, 2012.

Asymmetries in Effects of Monetary Policy and their Link to Labour Market – Application of TVP-VARs to Three European Countries

Anna Dobešová [1], David Hampel [2]

Abstract. The aim of this paper is to investigate asymmetries in effects of monetary policy and their link to labour market. For this purpose we employ Time-Varying Parameter VAR model with stochastic volatility that is able to capture possible changes in transmission mechanism and in underlying structure of economy over time. Markov chain Monte Carlo method is used for estimation. For detection of monetary policy effects we include three variables to the models: interest rate, inflation and unemployment. The Czech Republic, Germany and Slovakia are selected for the analysis. Main economic theory believes that wage rigidities are significant factor inflicting output gap. Therefore, we investigate how different responses to monetary policy operations over time are linked with changes in labour costs in observed countries. We are interested mainly in their connection with the current crisis. Our results prove asymmetric transmission over time and over countries. Significant increase of stochastic volatility and rise in responses is detected in time of structural break in 2008. Size of monetary policy effect is connected with labour costs, but this connection is not symmetric among countries.

Keywords: labour costs, monetary policy, time-varying parameters, vector autoregression.

JEL Classification: C32, E52
AMS Classification: 62P20, 62H12

1 Introduction

Central banks affect nominal and real economic variables by their operations with interest rates. New Keynesian monetary economics believes that ideal economy without nominal rigidities and minimal output gap can be reach only through stable price level. Medium-term lags in transmission mechanism are assumed since the days of monetarism [9]. Hence central banks try to achieve stable inflation rate in medium term. Thus, labour market and unemployment should not be affected, which is connected with modern Phillips curve relation.

Vector autoregressive (VAR) model is an often used tool for study of transmission mechanism. A lot of papers however found significant changes in monetary policy efficiency over time, e.g. [5], [14], [15]. Time-Varying Parameter VAR (TVP-VAR) with stochastic volatility is able to capture possible changes in variables relationship and in heteroskedasticity of innovations, too. It is used for modelling of structural breaks and for gradual changes in economy. Since Primiceri [18] has proposed TVP-VAR model, it is employed for historical analysis and for forecasting [8], [11], [13], [16].

Therefore, we want to use TVP-VAR models for our analysis. The year 2008 was connected with the onset of the economic crisis, hence we assume changes in models and especially in stochastic volatility after this date. Also other events, e.g. an entrance to the monetary union, can have significant impact to a transmission changes. We will able to capture cross-time asymmetries and potential cross-country asymmetries in these models. But information, that some non-linearity has occurred, is not sufficient. We want to know, what drives these non-linearities. There is part of literature focused on a detection of determinants of monetary transmission. Traditional approach explains cross-country asymmetries by correlation of one impulse-response characteristic with an average of possible factor for some period [3], [4]. This approach can be criticized for an abstracting from time-changes. Therefore, other methods are developed, for example Panel Conditionally Homogenous VAR [10] tries to capture factors time-variation by adjusting of coefficient matrix on some factor. It may be consider too theoretical approach with an unclear link to reality.

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, Czech Republic, e-mail: xdobeso5@node.mendelu.cz.
[2] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, Czech Republic, e-mail: qqhampel@mendelu.cz.

We want to link advantages of both approaches: clear interpretability of classical approach and capturing of changes of factors over time. Therefore, we suggest method in which we estimate trajectories of impulse-responses that imply variations of monetary transmission efficiency over time and then we make a regression with labour costs - possible monetary policy determinant - and we will observe its significance. We are focused on a link with labour market, because its situation has a key importance for efficiency of demand shock according to the main macroeconomic theory. We include three variables to the model: short-term interest rate, inflation and unemployment. We make the regression with unit labour costs, which can be consider as an opposite of labour costs rigidities. We can determine not only its influence to the monetary transmission but also to the relation of Phillips curve because of included variables. We have selected three different countries to the analysis, thus we will able to check the accuracy of the proposed method: the Czech Republic without euro, Germany, that has entered to European Monetary Union (EMU) in 1999, and Slovakia, that has been a member of EMU since 2009.

The aim of this contribution is to investigate asymmetries in monetary transmission and to explore their link to labour market. The rest of the paper is organized as follows: section 2 introduces data and methodology, section 3 dissects results and the last section concludes.

2 Data and Methods

We use three variables to TVP-VAR model: 3-months interbank interest rate, Harmonised Index of Consumer Prices (HICP) measured as a change compared to corresponding period of the previous year, and unemployment – monthly average. We include monthly observation from January 1999 to January 2014. We employ data in logarithmic differences, thus a stationarity of time series is ensured. Stationarity in context of TVP-VAR model does not have to be considered as necessary, because coefficients can freely move [19]. But time-variant coefficients can outweigh data and cause obscure relations, hence stationary is preferable [16]. We make a regression with unit labour costs measured as a change compared to corresponding period of the previous year. Therefore we model real changes not trends. Labour costs are available at quarterly frequency from 2001/Q01 to 2013/Q04, hence we convert impulse-response trajectories before the regression. Database of Eurostat is a source of our data.

TVP-VAR model for $t = 1, 2, \ldots, n$ can be defined as:

$$y_t = x_t \beta + z_t \alpha_t + \varepsilon_t, \tag{1}$$

where y_t is a scalar of responses, x_t, $(k \times 1)$, and z_t, $(p \times 1)$, are vectors of covariates, β is a $(k \times 1)$ vector of constant parameters, α_t is a $(p \times 1)$ vector of time-varying parameters and $\varepsilon_t \approx N(0, \sigma_t^2)$. Parameters $\alpha_{t+1} = \alpha_t + u_t$ follow random walk process for $t = 0, \ldots, n-1$, where $u_t \approx N(0, \Sigma)$, $\alpha_0 = 0$, $h_0 = 0$ and $\gamma > 0$. In $\sigma_t^2 = \gamma \exp(h_t)$ a stochastic volatility is $h_{t+1} = \phi h_t + \eta_t$, where $\eta_t \approx N(0, \sigma_t^2)$ for $t = 0, \ldots, n-1$. More detailed definition of TVP-VAR model is given in [16] or [18].

We employ Markov Chain Monte Carlo (MCMC) method for estimates of model parameters. We follow algorithm of Nakajima [16], for detailed explanation see his paper. We execute the estimations in 20 000 iterations and include two lags. The resulting models are sufficiently robust and there are no autocorrelations. All calculations were carried out in computational system Matlab R2011b.

3 Results and Discussion

Estimations of stochastic volatility of TVP-VAR models are in *Figure 1*. We can see an increase of volatility of inflation during break in 2008 in all countries and in 2004 in the Czech Republic, when it has entered to the European Union. There is a growth in stochastic volatility of unemployment around 2008 in the Czech Republic and Germany but with different trends and there is an increasing trend in Slovakia. These results are in accordance with the economic reality.

Impulse responses trajectories are illustrated in *Figure 2* (responses for nine-month horizon are chosen as a representative, because trends in other horizon are similar, but with different measure). There is a price puzzle effects, because response of inflation to positive shock in interest rate are positive and responses of unemployment are negative in ninth month. It is an opposite of the theory presumption. But this is consistent with the data that initial negative response for example of inflation to positive policy shock is displaced and transfigured to the positive one. Another fact is that we don't model data in absolute size but differences that means real relations. Furthermore, we can see in *Figure 2* an increase in monetary transmission effectiveness around structural break

in all countries. However, trends differ among states: a rise in monetary policy effects is visible in the Czech Republic, a decrease is in Germany and similar before- and during-crisis values are in Slovakia. It means that there are cross-country asymmetries in monetary transmission. Right panel in *Figure 2* determines Phillips curve relation: roughly inflation has negative impact to unemployment, but the link is unsteady, in some periods it is positive, too.

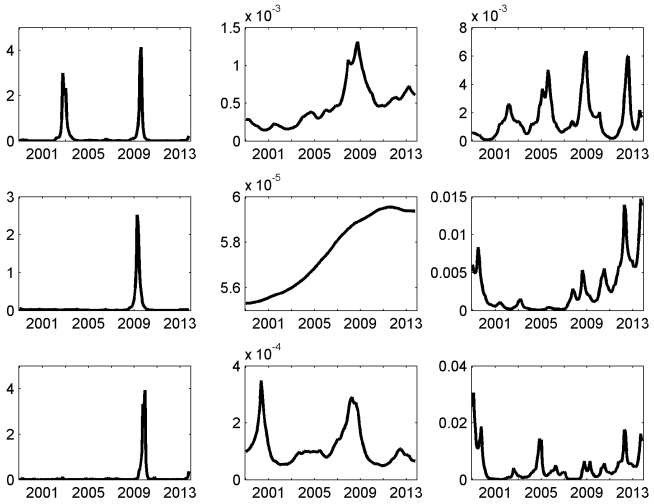


Figure 1 Stochastic volatility: left panel inflation, middle panel unemployment, right panel interest rate; first row Czech Republic, second row Germany, third row Slovakia

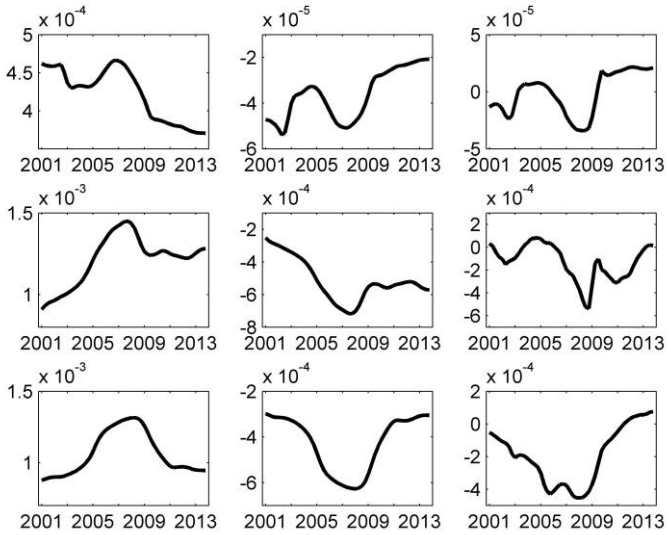


Figure 2 Trajectories of impulse responses for nine months horizon: left panel interest rate → inflation, middle panel interest rate → unemployment, right panel inflation → unemployment; first row Czech Republic, second row Germany, third row Slovakia

Development of labour costs is in *Figure 3*. The Czech Republic is characterized by decreasing increments in unit labour costs from nearly 20 % in 2001 to negative indices in 2013. Fluctuations of labour costs are more random in Germany and it circulates in narrower band than in the Czech Republic. Lower increments before crisis in Germany can be linked with the Hartz reforms of labour market there (from more details see [1]). Further, an increase during the start of the crisis can be related with packages of stimulus measures (so called Konjunkturpaket, see e.g. [2]). There is an observable decrease in labour costs changes in Slovakia. There are two breaks in Slovak series that divides the series into three periods. Slovak monetary policy had operated under the implicit inflation targeting until 2004. It was changed in 2005 to explicit targeting and National Bank of Slovakia had tried to meet Maastricht inflation and exchange rate criteria – this period was connected with grad-

ual labour costs decrease. Slovak Republic has entered to the EMU in 2009 and development of costs seems to be more random from that year, similar to the German one.

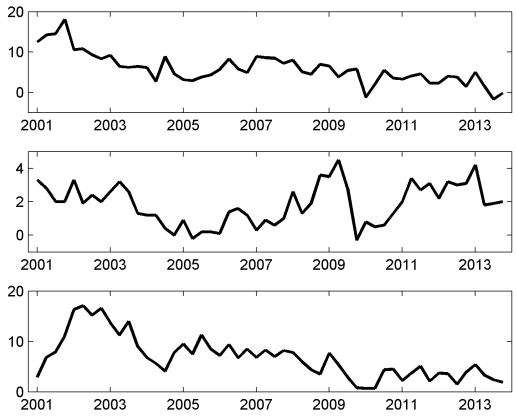


Figure 3 Labour costs: first row Czech Republic, second row Germany, third row Slovakia, source: Eurostat Labour cost index dataset.

In a further step we built regression models in which trajectories of impulse responses for individual monetary policy horizons are dependent variable and independent variables are labour costs and constant. The estimated values of the slope parameters of regression lines and p-values of their statistical significance are listed in *Table* **1** - *Table 3*. Results for the Czech Republic are in *Table 1*. Labour costs are significant regressor of monetary policy power for all periods in the Czech Republic. Calculated signs of estimated slopes are in accordance with theoretical expectations. For example, reduction of interest rate should cause positive demand shock. According to the contractual approach to labour market, long-term employment contracts cause rigidities, which imply slower adaptations of aggregate demand. Therefore, demand shock is able to affects real variables, i.e. output and unemployment, in short-term. Aggregate supply along with labour market will adapt in long-term and economy will return to the potential output. It means that long-terms effect of demands shock should be only nominal, i.e. inflationary. This is, of course, only model description, economic reality is more complex, but results for the Czech Republic complies with this theory. Changes in labour costs are basically the opposite of wage rigidities. Thus changes in labour costs allow faster adaptation of the aggregate supply during demand shock and more significant effects to inflation and less influence to unemployment are displayed (first two rows in *Table 1*). Last row in the table shows how is Phillips curve relation influenced by labour cost changes. Unemployment should be decreased after an increase in inflation. But if at the same time labour costs are increased, employers will not recruit so many employees. Therefore, labour cost has counter impact to Phillips curve, as it is expressed by negative slope in the *Table 1*.

We can mention that described relations are probably two-way and monetary transmission should inflict labour costs, too. But labour costs are considered to be more rigid in this case.

	3M	**6M**	**9M**	**12M**	**15M**	**18M**
i → p	7.0841e-06 (0.000)	1.1567e-05 (0.000)	6.22e-06 (0.000)	2.4229e-06 (0.000)	8.7061e-07 (0.000)	2.9374e-07 (0.000)
i → u	-1.2859e-05 (0.000)	-5.3597e-06 (0.000)	-2.0676e-06 (0.000)	-6.8976e-07 (0.000)	-2.1841e-07 (0.000)	-6.5878e-08 (0.000)
p → u	-4.288e-05 (0.000)	-9.9336e-06 (0.000)	-2.8265e-06 (0.000)	-7.3886e-07 (0.000)	-2.0001e-07 (0.000)	-5.3381e-08 (0.000)

Table 1 The Czech Republic, estimates of slope parameter with p-value in parentheses from regression model in which trajectories of impulse responses for individual horizons are dependent variable and labour costs with constant are independent variable.

Results for regression of impulse-response trajectories with labour costs in Germany are in *Table 2*. Labour costs also significantly influence monetary transmission, mainly in longer horizon, although p-values are not as low as in the Czech Republic. But signs of slope parameters differs from that Czech ones. It can be caused by a specific labour market situation in Germany. High saving rate, which has increased during the crisis even despite of the stagnation of real wages, is typical for Germany [17]. Thus, change in the interest rate is not able to cause so

large demand shock. Furthermore, we can assume more rigid aggregate supply there. Agenda 2010 has brought a differentiation of labour market and a strong protection of employees on full-time. Stimulus measures Konjunkturpaket [2] and Kurzarbeit, instrument of employment policy with which it is better for enterprises not to dismiss employees (see e.g. [6]), also imply greater rigidity of the aggregate supply. All this leads to the atypical behaviour of German labour market. Therefore, labour costs works in the opposite direction than in the Czech Republic. Monetary policy effects are more real and less inflationary when they are driven by labour costs. Costs have a significant influence to the Phillips curve only in 6-month horizon (third row in *Table 2*), but their sign is theoretically correct. This phenomenon can be caused by the labour market reforms, too.

	3M	**6M**	**9M**	**12M**	**15M**	**18M**
i → p	-1.1263e-05 (0.574)	-5.1331e-05 (0.005)	-3.8815e-05 (0.027)	-3.7908e-05 (0.007)	-3.0565e-05 (0.003)	-2.3598e-05 (0.001)
i → u	2.9108e-05 (0.021)	2.9023e-05 (0.091)	3.3367e-05 (0.026)	2.9451e-05 (0.011)	2.4087e-05 (0.003)	1.8488e-05 (0.001)
p → u	8.4871e-05 (0.127)	-7.2636e-05 (0.015)	-2.5661e-05 (0.162)	-1.8702e-05 (0.121)	-1.104e-05 (0.153)	-6.8536e-06 (0.167)

Table 2 Germany, estimates of slope parameter with p-value in parentheses from regression model in which trajectories of impulse responses for individual horizons are dependent variable and labour costs with constant are independent variable.

Table 3 shows results of labour costs with the monetary transmission strength in Slovakia. Labour costs are significant here only in 3-month horizon here. This is probably due to the important changes that occurred in the period in Slovakia: National Bank of Slovakia has transformed monetary regime and has entered to the monetary union. Contemporary effectiveness of the ECB monetary policy to Slovak variables is quite different from the former one here [15]. Influence of inflation to unemployment is however significantly inflicted by labour costs in the same direction as in the previous countries.

	3M	**6M**	**9M**	**12M**	**15M**	**18M**
i → p	-2.2636e-05 (0.000)	-1.2026e-05 (0.065)	-4.3521e-06 (0.392)	-9.2093e-07 (0.780)	2.6406e-07 (0.892)	5.1672e-07 (0.638)
i → u	2.063e-05 (0.006)	4.1905e-06 (0.513)	-2.8326e-07 (0.946)	-1.128e-06 (0.641)	-1.0168e-06 (0.450)	-7.2886e-07 (0.318)
p → u	-5.5081e-05 (0.012)	-2.6241e-05 (0.022)	-1.2263e-05 (0.033)	-5.9018e-06 (0.042)	-2.9085e-06 (0.047)	-1.4572e-06 (0.048)

Table 3 Slovakia, estimates of slope parameter with p-value in parentheses from regression model in which trajectories of impulse responses for individual horizons are dependent variable and labour costs with constant are independent variable.

Our approach provides results in conformity with the reality. Our findings shows asymmetries in monetary transmission over time, especially more persistent monetary policy shocks during the crisis. It is in accordance with results of Franta et al. [8] for the Czech Republic, although their impulse responses have opposite direction then ours. This is perhaps due to the fact that they work with non-stationary series and thus model mainly trends. We find out also cross-country asymmetries with significant correlation with labour costs. We prove strong interdependency of inflation and labour costs in the Czech Republic. Kai et al. [12] reached similar results for Germany in DSGE model: labour market settings can inflict monetary transmission and wage rigidities have an importance for inflation persistence and for unemployment fluctuation. Eickmeier and Pijnenburg [7] also concluded that unit labour cost has significant influence to inflation in Phillips curve. For more detailed results we will extend our future research to other countries and other factors.

4 Conclusions

The situation on labour market, especially rigidities caused by long-term work contracts, is one of the main reasons causing the output gap. Therefore, central banks try to achieve stable inflation rate and thus optimize product and minimize wage rigidities. Hence, this paper is focused on the link of monetary transmission with the labour market. We calculated changes in impacts of monetary policy shocks to inflation and unemployment and variation in dynamic Phillips curve relation. We concluded that stochastic volatility and persistence of shocks was greatest during structural break in 2008. But trends in monetary transmission differ over countries. We proved significant influences of labour costs to monetary transmission. Labour costs cause more nominal effects

of interest rate shocks in the Czech Republic. Contrary to that, labour costs in Germany intensify real impacts of monetary policy operations, which can be caused by reforms of German labour market. Situation in Slovakia is not so clear probably due to its entry into EMU and other changes in the Slovak Republic. Furthermore, changes in labour costs negatively affect Phillips curve relation in all countries. Our approach is appropriate for evaluating factors, which drive monetary transmission efficiency.

References

[1] Akyol, M., Neugart, M., Pichler, S.: Were the Hartz Reforms Responsible for the Improved Performance of the German Labour Market? *Economic Affairs* **33** (2013), 34-47.

[2] Barabas, G., Döhrn, R., Gebhardt, H., Schmidt, T.: Was bringt das Konjunkturpaket II? *Wirtschaftsdienst* **89** (2009), 128-132.

[3] Carlino, G., DeFina, R.: The Differential Regional Effects of Monetary Policy: Evidence from the U.S. States. *Journal of Regional Science* **39** (1999), 339-358.

[4] Dedola, L. Lippi, F.: The Monetary Transmission Mechanism: Evidence from the Industries of Five OECD Countries. *European Economic Review* **49** (2005), 1543–1569.

[5] Égert, B., CrespoCuaresma, J., Reininger, T.: Interest Rate Pass-through in Central and Eastern Europe: Reborn From Ashes Merely to Pass Away? *Journal of Policy Modeling* **29** (2007), 209-225.

[6] Eichhorst, W., Marx, P.: Kurzarbeit: Sinnvoller Konjunkturpuffer oder verlängertes Arbeitslosengeld? *Wirtschaftsdienst* **89**, 322-328.

[7] Eickmeier, S. and Pijnenburg, K.: The Global Dimension of Inflation - Evidence from Factor-Augmented Phillips Curves. *Oxford Bulletin of Economics and Statistics* **75** (2013), 103–122.

[8] Franta, M., Horváth, R., Rusnák, M.: *Evaluating Changes in Monetary Transmission Mechanism in the Czech Republic*. ČNB Working Paper Series. Czech National Bank, Prague, 2011.

[9] Friedman, M.: The Quantity Theory of Money, A Restatement. In: Friedman, M. (ed.): *Studies in the Quantity Theory of Money*, University of Chicago Press, Chicago, 1956.

[10] Georgiadis, G.: *Towards an Explanation of Cross-country Asymmetries in Monetary Transmission*. Deutsche Bundesbank Discussion Papers, Deutsche Bundesbank, Frankfurt am Main, **7** (2012).

[11] Koop, G., Korobilis, D.: Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. *Foundations and Trends(R) in Econometrics* **3** (2010), 267-358.

[12] Kai, C., Kuester, K., Linzert, T.: *The Impact of Labor Markets on the Transmission of Monetary Policy in an Estimated DSGE Model*. IZA Discussion Papers, Institute for the Study of Labor, Bohn, 2005.

[13] Kumar, M.: A time-varying parameter vector autoregression model for forecasting emerging market exchange rates. *International Journal of Economic Sciences and Applied Research*, **3** (2010), 21-39.

[14] Lyziak, T., Przystupa, J., Stanislawska, E., Wróbel, E.: Monetary Policy Transmission Disturbances during the Financial Crisis: A Case of an Emerging Market Economy. *Eastern European Economics* **49** (2011), 75-96.

[15] Myšková, K., Dobešová, A.: Comparing monetary policy before and during crisis for Visegrád group countries. In: *Proceedings of the 31st International Conference Mathematical Methods in Economics 2013* (Vojáčková, H., ed). College of Polytechnics Jihlava, Jihlava, 2013, 649-654.

[16] Nakajima, J.: Time-Varying Parameter VAR Model with Stochastic Volatility: An Overview of Methodology and Empirical Applications. *Monetary and Economic Studies* **30** (2011), 107–142.

[17] OECD: *Economic Surveys: Germany 2010*. OECD Publishing, Paris, 2010.

[18] Primiceri, G.: Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies* **72** (2005), 821–852.

[19] Sims, C. A., Stock, J. H., Watson, M. W.: Inference in Linear Time Series Models with Some Unit Roots. *Econometrica* **58** (1990), 113–44.

Categorical data analysis from Lean Company research

Radim Dolák[1], Jan Górecki[2], Lukáš Slechan[1,2], Michael Kubát[1,2]

Abstract. This article deals with dependencies among variables in data obtained from a questionnaire survey which focused on the issue of lean manufacturing and lean logistics. The main hypothesis of this paper is that the implementation of lean manufacturing and lean logistics in selected companies is dependent on the size of a company. Attention will be paid to the question of whether the implementation of lean manufacturing and lean logistics depends on the branch of industry such as engineering, electrotechnical industry or chemical industry. Source data were obtained using a questionnaire survey of manufacturing companies in the Czech Republic. The data are divided into four categories by the size of company: a micro company, a small company, a medium company and a large enterprise. Specialized statistical software IBM SPSS Statistics is used for the analysis of categorical data. There will be analyzed selected forms of dependencies based on data from Lean Company research.

Keywords: categorical data analysis, lean company research, SPSS.

JEL Classification: C14
AMS Classification: 62H17

1 Introduction

This article deals with dependencies between data obtained from the questionnaire survey, which focused on the issue of lean manufacturing and lean logistics. The main hypothesis of this paper is that the implementation of lean manufacturing and lean logistics is dependent on the size of the company. Attention will be paid to the question of whether implementation of lean manufacturing and lean logistics depends on the branch of industry such as engineering, electrotechnical industry or chemical industry. The main objective of data analysis is to verify the validity of the hypothesis and also try to discover additional dependencies that are not obvious at first sight.

Manufacturers around the world are now trying to embrace lean production. Lean production uses less of everything compared with mass production - half the human effort in the factory, half the manufacturing space, half the investment in tools, half the engineering hours to develop a new product in half the time [8]. Principles and methods of lean production were also elaborated on logistics. It is necessary to reduce unnecessary waste in manufacturing and logistics to achieve a competitive advantage in conditions of current global competition. We can find information about Lean company concept in [5, 8]. The analysis of data from surveys is described for example in [7]. We can use a variety of methods and approaches for data analysis such as categorical data analysis [1, 2, 6] or for example methods of data-mining [3].

The paper is structured as a follows: section "Categorical data analysis" is theoretical part about categorical data analysis, section "Lean Company research" describes structure and process of our research and section "Categorical data analysis in SPSS" deals with results of dependency analysis of our data from Lean Company research.

2 Categorical data analysis

What is a categorical variable? A categorical variable has a measurement scale consisting of a set of categories [2]. For example, implementation of lean manufacturing concept should be measured as fully implemented, partially implemented, will be implemented or not will be implemented.

There are two statistics that are used to determine whether or not two variables are independent of each other: The Pearson χ^2 and likehood-ratio G^2.These statistics provide two separate test statistics, but they share many

[1] Silesian university in Opava: School of Business Administration in Karvina, Department of Informatics, Univerzitní náměstí 1934/3, Karviná, Czech Republic, dolak@opf.slu.cz.
[2] Silesian university in Opava: School of Business Administration in Karvina, Department of Informatics, Univerzitní náměstí 1934/3, Karviná, Czech Republic, gorecki@opf.slu.cz.

properties and usually provide the same conclusions. When H_0 is true and the expected frequencies are large, the two statistics have the same chi-squared distribution, and their numerical values are similar [1].

Pearson Statistic and Chi-Squared Distribution

We consider according to Agresti [1] the null hypothesis H_0 that cell probabilities equal certain fixed values (π_{ij}). For a sample of size n with cells counts (n_{ij}), the values ($\mu_{ij} = n\pi_{ij}$) are expected frequencies. They represent the values of the expectations ($E(n_{ij})$) when H_0 is true.

Pearsons chi-squared statistic for testing H_0 is defined as the following equation

$$\chi^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \tag{1}$$

This statistics takes its minimum value of zero when all $n_{ij=}\ \mu_{ij}$. For a fixed sample size, greater differences (n_{ij}- μ_{ij}) produce larger χ^2 values and stronger evidence against H_0 [1].

Likehood ratio statistic

The Likehood function is according to Agresti [1] the probability of the data, viewed as a function of the parameter once the data are observed. For two-way contingency tables with likehood function based on the multinomial distribution, the likehood-ratio statistic simplifies to

$$G^2 = 2\sum n_{ij} \log\left(\frac{n_{ij}}{\mu_{ij}}\right) \tag{2}$$

Software for categorical data analysis

Statistical methods and techniques for categorical data analysis have undergone rapid development in the past 25 years. Their application in applied research have become commonplace in recent years, due in large part to the availability of commercial software and inexpensive computing [6]. There are many ways how to analyze categorical data. We can use for example widespread spreadsheet Microsoft Excel or specialized statistical software as IBM SPSS Statistics, SAS or STATA. There will be used specialized statistical software IBM SPSS Statistics for the analysis of categorical data and their dependencies in this paper. One of its strong points is that SPSS is capable of performing almost all common types of analysis. This makes the software particularly suitable for analyzing data obtained from questionnaires [4].

3 Lean company research

Our data were obtained using a questionnaire survey of manufacturing companies in the Czech Republic. We sent 3500 questionnaires to various companies. We obtained data from 101 different companies which are analyzed in next section "Categorical data analysis in SPSS".

There are many non-metric data in our Lean company research. Nonmetric data is a categorical measurement and is expressed not in terms of numbers but rather by means of a natural language description. It is often known as categorical data. These data can be measured on two different scales that is, nominal and ordinal [9].

The data could be separated into many categories according to:

- the company size (number of employees);
- the branch of industry;
- the distribution by region;
- the knowledge of lean manufacturing;
- the knowledge of logistics;
- the training of lean manufacturing;
- the training of lean logistics;
- the level of implementation of lean manufacturing concept;
- the level of implementation of lean logistics concept;
- the using ICT to support lean company concept;
- the using expert for lean manufacturing;
- the using expert for lean logistics.

Companies could be divided according their size as: a micro company, a small company, a medium company and a large enterprise. There are 3 branches of industry: electro-technical, chemical, mechanical engineering. For many categories such as the knowledge of lean manufacturing and the knowledge of lean logistics, the training of lean manufacturing and lean logistics are possible only positive or negative answers yes or no. Level of implementation of lean manufacturing and lean logistics concept could be one of following: fully implemented, partially implemented, will be implemented, will not be implemented.

4 Categorical data analysis in SPSS

Here we describe the analysis of the data acquired from our Lean company research. We use IMB SPSS Statistics software to perform the analysis in this paper. We focus on categorical data analysis divided into the most important categories to sort data from our research: the company size (number of employees), the branch of industry and finally dependencies between knowledge and trainings of lean company concepts (manufacturing and logistics). We show basic results from SPSS with a commentary of an example. Other combinations of categorical data and its independence or dependence between the variables is shown in tables 1, 2, 3. Every figure and table shows the most important information from SPSS analysis such as: names of the variables between which we seek for relations, then results of Pearson Chi-Square and Asymp. Significance (2-sided) and finally there is a conclusion about independence or dependence between the variables.

4.1 Analysis of dependencies between the company size and other variables

The following Figure 1 shows crosstabulation with counts of categories separated according to the *company size (employees)* and variable the *knowledge of lean manufacturing concept*. Figure 2 then shows output from SPSS with results of a dependency analysis for variables the *company size (employees)* and the *knowledge of lean manufacturing concept*.

		knowledge of lean manufacturing		
		YES	NO	Total
employees	1-99 employees	12	26	38
	100-499 employees	31	8	39
	500 and more employees	22	2	24
Total		65	36	101

Figure 1 Crosstabulation: company size (employees) * knowledge of lean manufacturing concept

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	29,493[a]	2	,000
Likelihood Ratio	30,825	2	,000
N of Valid Cases	101		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 8,55.

Figure 2 Chi-Square Tests: company size (employees) * knowledge of lean manufacturing concept

The first line (Pearson Chi-Square) on Figure 2 refers to the Pearson chi-square statistics from equation (1). The second line (Likehood Ratio) from equation (2) is likelihood ratio and the third line shows the total sample size. For both statistics is in the first column (Value) indicate the counted value statistics, in the second column (df) is degrees of freedom and the third column (Asymp. Sig. (2-sided)) shows the minimum level of significance of the test and rejects or accepts the hypothesis H_0 (independence between the variables). Because the minimum level of significance is less than 0,05 we reject the hypothesis H_0 of independence and accept alterna-

tive hypothesis H_A about dependence between variables. There are two important rules for Chi-squared test of independence that must be respected. A prerequisite for the use of this test is that the expected frequency in the individual cells counts (n_{ij}) did not fall below 5% in at least 80% cells and the remaining cells have occurred at least value 1 [7]. We can see in Figure 2 that 0 cells (0,0%) have expected count less than 5 and the minimum expected count is 8,55.

Previous Figure 2 was describing dependency between the company size (employees) and the knowledge of lean manufacturing concept. We also analyzed other pair wise dependencies in other pairs of the variables. The results of these analyzes can be seen in the next Table 1.

Relation between the variables	Pearson Chi-Square	Asymp. Sig. (2-sided)	Independence between the variables
employees * knowledge of lean manufacturing	29,493	,000	NO
employees * knowledge of lean logistics	15,428	,000	NO
employees * training of lean manufacturing	27,894	,000	NO
employees * training of lean logistics	17,058	,000	NO
employees * expert for lean manufacturing	10,089	,006	NO
employees * expert for lean logistics	6,316	,043	NO

Table 1 Independence between the company size (employees) and other variables: Chi-Square test in SPSS

We can see in Table 1 very strong dependency between company size (employees) and variables such as the knowledge of lean manufacturing, the knowledge of lean logistics, the training of lean manufacturing, the training of lean logistics.

The main hypothesis of this paper about dependency between the size of a company and implementation of lean manufacturing and lean logistics was confirmed. We used the 0,05 probability level as our critical value. We can see that Asymp. Sig. (2-sided) value is for both 0,000 and it is less than the 0,05 value so we reject H_0 about independence of variables and we except alternative hypothesis H_A about dependence between variables.

4.2 Analysis of dependencies between the branch of industry and other variables

The next Figures 3 and 4 shows results of analyze of independency between the *branch of industry* and the *knowledge of lean manufacturing* concept. We can see crosstabulation that shows us the division into categories by the *branch of industry* such as electrotechnical, chemical or mechanical engineering with its counts of the *knowledge of lean manufacturing concept.*

		knowledge of lean manufacturing		
		YES	NO	Total
branch of industry	Electrotechnical	9	11	20
	Chemical	8	2	10
	Mechanical engineering	48	23	71
Total		65	36	101

Figure 3 Crosstabulation: branch of industry * knowledge of lean manufacturing concept

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4,660[a]	2	,097
Likelihood Ratio	4,605	2	,100
N of Valid Cases	101		

a. 1 cells (16,7%) have expected count less than 5. The minimum expected count is 3,56.

Figure 4 Chi-Square Tests: branch of industry * knowledge of lean manufacturing concept

Previous Figure 4 was analyzing dependence between the *branch of industry* and the *knowledge of lean manufacturing concept*. We analyzed also more combinations of dependencies between other variables. We can see results from SPSS in next Table 2.

Relation between the variables	Pearson Chi-Square	Asymp. Sig. (2-sided)	Independence between the variables
branch of industry * knowledge of lean manufacturing	6,660	,097	YES
branch of industry * knowledge of lean logistics	7,492	,024	NO
branch of industry * training of lean manufacturing	2,542	,280	YES
branch of industry * training of lean logistics	1,990	,370	YES
branch of industry * expert for lean manufacturing	1,452	,484	YES
branch of industry * expert for lean logistics	1,229	,541	YES

Table 2 Independence between the branch of industry and other variables: Chi-Square test in SPSS

We can see that there is pair wise independency between branch of industry and variable such as knowledge of lean manufacturing, training of lean manufacturing, training of lean logistics, expert for lean manufacturing and expert for lean logistics. There is dependence only between branch of industry and knowledge of lean logistics concept.

We found answer for question if exists dependency between branch of industry and implementation of lean manufacturing and lean logistics. We used the 0,05 probability level as our critical value. We can see in Table 2 that Asymp. Sig. (2-sided) value for the knowledge of lean manufacturing is 0,097 and it is more than the 0,05 value so we accept H_0 about independence between the branch of industry and the knowledge of lean manufacturing. We can see in Table 2 that Asymp. Sig. (2-sided) value for the knowledge of lean logistics is 0,024 and it is less than the 0,05 value so we reject H_0 about independence of the branch of industry on the knowledge of lean logistics.

4.3 Analysis of dependencies between knowledge and training of lean company

The last analysis investigates dependencies between the combinations of variables such as: the knowledge of lean manufacturing, the knowledge of lean logistics, the training of lean manufacturing and the training of lean logistics. We can see results from SPSS in next Table 3.

Relation between the variables	Pearson Chi-Square	Asymp. Sig. (2-sided)	Independence between the variables
knowledge of lean manufacturing * knowledge of lean logistics	64,270	,000	NO
training of lean manufacturing * training of lean logistics	58,538	,000	NO
knowledge of lean manufacturing * training of lean manufacturing	57,057	,000	NO
knowledge of lean logistics * training of lean logistics	51,617	,000	NO

Table 3 Independence between the variables: Chi-Square test in SPSS

We see very strong dependencies in all of the pairs of the variables: the knowledge of lean manufacturing and the knowledge of lean logistics, the training of lean manufacturing and the training of lean logistics, the knowledge of lean manufacturing and the training of lean manufacturing and finally between the knowledge of lean logistics and the training of lean logistics.

5 Conclusion

This paper describes an analysis of data acquired from survey of manufacturing companies in the Czech Republic. We have analyzed pair wise dependencies in pairs of variables using two selected statistics: Pearson chi-squared χ^2 and Likehood ratio statistic G^2. We used specialized statistical software IBM SPSS Statistics to get results about dependencies between variables. The main hypothesis of this paper is that the implementation of lean manufacturing and lean logistics is dependent on the size of the company. We confirmed the main hypothesis based on results of Chi-Square test in SPSS in section "Analysis of dependencies between the company size and other variables". We have discovered also other dependencies in other pairs of variables which are related to our Lean Company research.

Acknowledgements

The work of Radim Dolák was funded by the project "Inovace studijních programů na Slezské univerzitě, Obchodně podnikatelské fakultě v Karviné, Reg. č. CZ.1.07/2.2.00/28.0017". The work of Jan Górecki was funded by the project SGS/21/2014 - Advanced methods for knowledge discovery from data and their application in expert systems, Czech Republic.

References

[1] Agresti, A.: *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Hoboken, 2007.
[2] Agresti, A.: *Categorical data analysis*. John Wiley & Sons, Hoboken, 2002.
[3] Góreccki, J. and Holeňa, M.: An Alternative Approach to the Structure Determination of Hierarchical Archimedean Copulas. In: *Proceedings of 31st International Conference Mathematical Methods in Economics.* Jihlava, 2013, 201 – 206.
[4] Huizingh, E.: *Applied Statistics with SPSS*. SAGE Publications Ltd, London, 2007.
[5] Košturiak, J., and Frolík, Z.: *Štíhlý a inovativní podnik*. Alfa Publishing: Praha, 2006.
[6] Powers, D. A., and Xie, Y.: *Statistical Methods for Categorical Data Analysis*. Emerald Group Publishing Limited, Bingley, 2008.
[7] Řezánková, H.: *Analýza dat z dotazníkových šetření*. Professional Publishing, Praha, 2011.
[8] Womack, J. P., Jones, D. T., and Roos, D.: *The Machine That Changed the World: The Story of Lean Production*. First Free Press Edition, New York, 2007.
[9] Verma, J.P.: *Data Analysis in Management with SPSS Software*, Springer, Heidelberg, 2013.

$M/E_n/1/m$ queuing system subject to catastrophes

Michal Dorda[1], Dušan Teichmann[2]

Abstract. The paper is focused on a finite single-server queuing system with a server subject to so-called catastrophes. The catastrophes represent a special type of failures that cause emptying the system (all the customers found in the system are flushed out of the system). When the server is up, the customers arrive according to a Poisson input process with the parameter $\lambda > 0$. When the server is under repair (down), all the arriving customers refuse to enter the system because they are not willing to wait. The customer's service time is assumed to be Erlang distributed with the shape parameter n and the scale parameter $n\mu$. That means the service process of each customer consists of n exponentially distributed phases and the customer is served after finishing the last phase of the service process. Times between catastrophes are exponentially distributed with the parameter η; times to repair are also exponentially distributed, but with the parameter ζ. In the paper we present a mathematical model of the studied queuing system that enables to numerically calculate steady-state probabilities. We discuss some performance measures of the system and some possible modifications of the studied queuing system. Some numerical results are presented in the paper.

Keywords: $M/E_n/1/m$, queuing, catastrophes, Matlab.

JEL Classification: C44
AMS Classification: 60K25

1 Introduction

Queuing systems with servers subject to catastrophes represent a class of the queuing systems subject to failures (sometimes so-called unreliable queuing systems) because the catastrophe is a special type of the failure. The concept of server's catastrophes is close to disasters. The usual concept of disasters consider that the system empties when a disaster occur in the system. That means all (or some) of the customers found in the system are lost (or rejected). When the server is under repair (sometimes it is assumed that the server is ready to start working immediately without any repair − so-called clearing models − see for example [1]) the arrival process of the customers is not usually interrupted. Often, the customers are impatient when the server is down − see for example [4] or [5]. In the case of the catastrophes it is usually assumed that the system empties as well but the arrival process of the customers is interrupted; that means no customer enters the system when the server is under repair − see [3].

Let us consider a real example from rail freight transportation. Freight trains usually begin and end their journey in marshalling yards. Each marshalling yard has two main functions − to classify and make up the freight trains. So-called inbound freight trains enter the marshalling yard and individual wagons (or groups of wagons) of the train are classified according to their transport relations in the marshalling yard. On the other hand when enough wagons have been gathered for a transport relation in the marshalling yard, a new outbound freight train can depart from the yard.

To carry out the tasks mentioned above the marshalling yards are equipped with a special facility − a hump − that is utilized for the classification process of the freight trains entering the yard. The hump represents a server of a queuing system that serves the inbound freight trains. The trains stopped on arrival tracks that form so-called reception sidings. It is clear that the number of the arrival tracks in the yard is finite and, therefore, the number of the freight trains that can wait in the yard is also finite. That is the reason why all of the queuing models mentioned in the previous paragraphs cannot be applied in this case − all of them assume that the queue length can be infinite. In the paper we present a mathematical model of a queuing system subject to catastrophes with a finite capacity of the input buffer which can be applied in the case we discussed. In the case of our example the catastrophe can be represented for example by a failure of a switch in a ladder or by derailing a vehicle in the area of the hump or the ladder. In such cases the classification process cannot be carried out via the hump; the inbound freight trains has to be classified in a different way. For example, the railway station "Ostrava hlavní nádraží"

[1] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 325 754, e-mail: michal.dorda@vsb.cz.
[2] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 324 575, e-mail: dusan.teichmann@vsb.cz.

has two hump yards – "Ostrava levé nádraží" and "Ostrava pravé nádraží". If it is not possible to classify the inbound trains for example in "Ostrava pravé nádraží" due to technical reasons, all the inbound trains waiting in "Ostrava pravé nádraží" have to be shunted to "Ostrava levé nádraží" where the trains can be consequently classified. That is why we assume emptying the system. Please note that a different way how to get the performance measures of the queuing system is based on simulation, in paper [2] we shortly described a simulation model based on coloured Petri nets that could be utilized after some minor modifications.

2 Mathematical model

Let us assume a single-server queuing system with the server subject to the catastrophes. The system consists of m places for the customers, where $m \geq 2$ − one place in the service and $m-1$ places in the queue. Customers waiting in the queue are served according to the ordinary FCFS discipline. From time to time the server suffers the catastrophic failure which empties the system. Moreover, all the customers arriving in the system when the server is down are not willing to wait and leave the system without being serviced − we consider that the customers are rejected.

Customers arrive in the system in a Poisson input process with the parameter $\lambda > 0$. Service times are considered to be Erlang distributed with the shape parameter $n \geq 2$ and the scale parameter $n\mu > 0$. Please note that the Erlang distributed service times with the parameters n and $n\mu$ consist of n exponentially distributed phases, all of them have the same parameter $n\mu$. The decomposition enables us to model the system as a Markov process.

Let times between the catastrophes be exponentially distributed with the parameter η. The parameter η is greater than 0 if the system is idle or busy and equal to 0 if the system is down; that means the catastrophes can occur only when the system is idle or busy. Times to repair are also exponentially distributed, but with the parameter $\zeta > 0$. We assume that the server is immediately started to be repaired after the catastrophes, and immediately starts to serve customers when repaired.

The server can be found in one of the following states:

- The server is idle − there is no customer in the system and the server is up. Let us denote the steady-state probability the server is idle as P_{idle}, the probability also expresses the ratio of time when the server is idle.
- The server is busy − there is at least one customer in the system and the server is serving one of them, other customers are waiting in the queue. Let us denote the corresponding steady-state probability P_{busy}, the probability corresponds to the ratio of time the server is busy. The probability P_{busy} is given by formula:

$$P_{busy} = \sum_{k=1}^{m} \sum_{p=0}^{n-1} P_{(k,p)}, \tag{1}$$

  where k = 1,..., m represents the number of the customers in the system and p = 0,..., $n-1$ expresses how many phases of the service of the customer which is being served the server has already finished. If the system is found in state (k,p), k customers are in the system and p phases of the service have been finished.
- The server is down − the server is down and under repair; there is no customer in the system. The steady-state probability the server is down is denoted as P_{down} and expresses the ratio of time the server is broken and therefore not able to serve the customers.

All the individual states (nodes of the state transition diagram) and possible transitions between them (arcs of the state transition diagram) are shown in Figure 1. Each arc is labelled by a corresponding transition rate. The meaning of the individual states was described in the previous text.

For the probabilities P_{idle}, P_{busy} and P_{down} the normalization condition of probability has to be satisfied:

$$P_{idle} + P_{busy} + P_{down} = 1. \tag{2}$$

It can be easily proved that it holds:

$$P_{down} = \frac{\frac{1}{\zeta}}{\frac{1}{\eta} + \frac{1}{\zeta}} = \frac{\eta}{\eta + \zeta} \tag{3}$$

and therefore:

$$P_{idle} + P_{busy} = 1 - P_{down} = \frac{\zeta}{\eta + \zeta}. \tag{4}$$

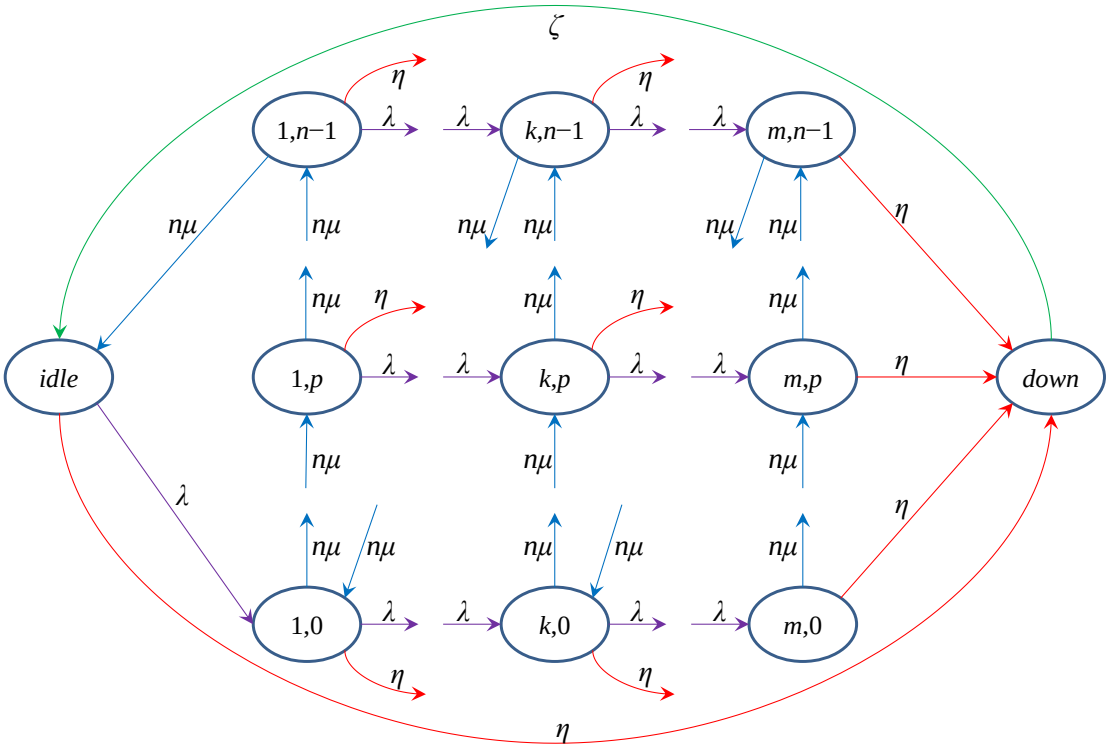


Figure 1 The state transition diagram

Now, using global balance principle (the flow out of each state has to be equal to the flow into the state) and the state transition diagram of the modelled system we can write down linear formulas for the steady-state probabilities of the individual states:

$$(\lambda + \eta) \cdot P_{idle} = n\mu \cdot P_{(1,n-1)} + \zeta \cdot P_{down}, \tag{5}$$

$$(\lambda + n\mu + \eta) \cdot P_{(1,0)} = \lambda \cdot P_{idle} + n\mu \cdot P_{(2,n-1)}, \tag{6}$$

$$(\lambda + n\mu + \eta) \cdot P_{(1,p)} = n\mu \cdot P_{(1,p-1)} \text{ for } p = 1,\ldots,n-1, \tag{7}$$

$$(\lambda + n\mu + \eta) \cdot P_{(k,0)} = \lambda \cdot P_{(k-1,0)} + n\mu \cdot P_{(k+1,n-1)} \text{ for } k = 2,\ldots,m-1, \tag{8}$$

$$(\lambda + n\mu + \eta) \cdot P_{(k,p)} = \lambda \cdot P_{(k-1,p)} + n\mu \cdot P_{(k,p-1)} \text{ for } k = 2,\ldots,m-1 \text{ and } p = 1,\ldots,n-1, \tag{9}$$

$$(n\mu + \eta) \cdot P_{(m,0)} = \lambda \cdot P_{(m-1,0)} \tag{10}$$

$$(n\mu + \eta) \cdot P_{(m,p)} = \lambda \cdot P_{(m-1,p)} + n\mu \cdot P_{(m,p-1)} \text{ for } p = 1,\ldots,n-1 \tag{11}$$

$$\zeta \cdot P_{down} = \eta \cdot P_{idle} + \eta \cdot \sum_{k=1}^{m} \sum_{p=0}^{n-1} P_{(k,p)} \tag{12}$$

Summing equations (5) and (12):

$$(\lambda + \eta) \cdot P_{idle} + \zeta \cdot P_{down} = n\mu \cdot P_{(1,n-1)} + \zeta \cdot P_{down} + \eta \cdot P_{idle} + \eta \cdot \sum_{k=1}^{m} \sum_{p=0}^{n-1} P_{(k,p)}. \tag{13}$$

After some algebraic manipulations we get from (13):

$$P_{idle} = \frac{n\mu}{\lambda} \cdot P_{(1,n-1)} + \frac{\eta}{\lambda} \cdot \sum_{k=1}^{m} \sum_{p=0}^{n-1} P_{(k,p)}. \tag{14}$$

Substituting (14) into equation (6) we get:

$$(\lambda + n\mu + \eta)\cdot P_{(1,0)} = \lambda \cdot \left[\frac{n\mu}{\lambda} \cdot P_{(1,n-1)} + \frac{\eta}{\lambda} \cdot \sum_{k=1}^{m} \sum_{p=0}^{n-1} P_{(k,p)} \right] + n\mu \cdot P_{(2,n-1)} . \tag{15}$$

After some algebraic manipulations with (15) we get:

$$(\lambda + n\mu)\cdot P_{(1,0)} = \eta \cdot \sum_{p=1}^{n-2} P_{(1,p)} + (n\mu + \eta)\cdot P_{(1,n-1)} + \eta \cdot \sum_{p=0}^{n-2} P_{(2,p)} + (n\mu + \eta)\cdot P_{(2,n-1)} + \eta \cdot \sum_{k=3}^{m} \sum_{p=0}^{n-1} P_{(k,p)} . \tag{16}$$

Equations (7) − (11) and (16) form an equation system with $m \cdot n$ unknown probabilities $P_{(k,p)}$, where $k = 1,\ldots,m$ and $p = 0,\ldots,n-1$. Now, in each equation (7) – (11) and (16) we subtract the left side of the equation to get zero on the left side. By this we get the equation system which can be generally rewritten in the following matrix form:

$$\begin{vmatrix} \mathbf{B_1} & \mathbf{C_1} & \mathbf{C_2} & \cdots & \cdots & \mathbf{C_2} \\ \mathbf{A} & \mathbf{B} & \mathbf{C} & & & \\ & \mathbf{A} & \mathbf{B} & \mathbf{C} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \mathbf{A} & \mathbf{B} & \mathbf{C} \\ & & & & \mathbf{A} & \mathbf{B_m} \end{vmatrix} \cdot \begin{vmatrix} P_{(1,0)} \\ P_{(1,1)} \\ \vdots \\ P_{(k,p)} \\ \vdots \\ P_{(m,n-1)} \end{vmatrix} = \mathbf{0} , \tag{17}$$

where all the sub-matrices $\mathbf{A}, \mathbf{B_1}, \mathbf{B}, \mathbf{B_m}, \mathbf{C_1}, \mathbf{C_2}, \mathbf{C}$ are $n \times n$. For example, for n=3 and $m \geq 3$ we get the following sub-matrices:

$$\mathbf{A} = \begin{vmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{vmatrix}, \ \mathbf{B_1} = \begin{vmatrix} -(\lambda + n\mu) & \eta & n\mu + \eta \\ n\mu & -(\lambda + n\mu + \eta) & 0 \\ 0 & n\mu & -(\lambda + n\mu + \eta) \end{vmatrix},$$

$$\mathbf{B} = \begin{vmatrix} -(\lambda + n\mu + \eta) & 0 & 0 \\ n\mu & -(\lambda + n\mu + \eta) & 0 \\ 0 & n\mu & -(\lambda + n\mu + \eta) \end{vmatrix}, \ \mathbf{B_m} = \begin{vmatrix} -(n\mu + \eta) & 0 & 0 \\ n\mu & -(n\mu + \eta) & 0 \\ 0 & n\mu & -(n\mu + \eta) \end{vmatrix},$$

$$\mathbf{C_1} = \begin{vmatrix} \eta & \eta & n\mu + \eta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}, \ \mathbf{C_2} = \begin{vmatrix} \eta & \eta & \eta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} \text{ and } \mathbf{C} = \begin{vmatrix} 0 & 0 & n\mu \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}.$$

Due to the fact that one equation is redundant in equation system (17), in order to get a unique solution of the equation system a row of the system can be omitted and the equation system has to be supplemented by the normalization equation of probability (2) into which we must substitute formulas (1), (3) and (14):

$$\frac{n\mu}{\lambda} \cdot P_{(1,n-1)} + \frac{\eta}{\lambda} \cdot \sum_{k=1}^{m} \sum_{p=0}^{n-1} P_{(k,p)} + \sum_{k=1}^{m} \sum_{p=0}^{n-1} P_{(k,p)} + \frac{\eta}{\eta + \zeta} = 1 . \tag{18}$$

After some algebraic manipulations we can write down the modified normalization equation (18) in the following form:

$$\left(\frac{\eta}{\lambda} + 1\right) \cdot \sum_{p=0}^{n-2} P_{(1,p)} + \left(\frac{n\mu + \eta}{\lambda} + 1\right) \cdot P_{(1,n-1)} + \left(\frac{\eta}{\lambda} + 1\right) \cdot \sum_{k=2}^{m} \sum_{p=0}^{n-1} P_{(k,p)} = 1 - \frac{\eta}{\eta + \zeta} . \tag{19}$$

Solving the modified equation system can be carried out numerically using suitable software; we utilized Matlab. After solving the system we get the stationary probabilities $P_{(k,p)}$, summing them we get P_{busy}. The probability P_{down} can be computed directly from formula (3) and the probability P_{idle} can be computed using formula (2) or (14).

2.1 Some performance measures

On the basis of the steady-state probabilities which have been numerically computed, we are able to compute several performance measures of the studied queuing system.

Due to the fact that the modelled system is equipped only with single server, the mean number of the customers in the service *ES* is equal to the utilization of the server ρ and these performance measures equal to the steady-state probability the server is busy. Therefore we can write:

$$ES = \rho = P_{busy} = \sum_{k=1}^{m}\sum_{p=0}^{n-1} P_{(k,p)} \,. \tag{20}$$

The mean number of the customers waiting in the queue can be computed using the individual probabilities $P_{(k,p)}$ according to the formula for the mean value of a discrete random variable, where the random variable is the number of the customers waiting in the queue. Hence we can write:

$$EL = \sum_{k=2}^{m}(k-1)\cdot\sum_{p=0}^{n-1} P_{(k,p)} \,. \tag{21}$$

For the mean number of the customers found in the system *EK* it has to hold:

$$EK = ES + EL = \sum_{k=1}^{m} k\cdot\sum_{p=0}^{n-1} P_{(k,p)} \,. \tag{22}$$

The customer is rejected upon the arrival when the system is full (that means there are *m* customers in the system) or when the server is down (the customer is not willing to wait and refuse to enter the system). The system is full with the probability:

$$P_{full} = P_m = \sum_{p=0}^{n-1} P_{(m,p)} \tag{23}$$

and the server is down with the probability P_{down}. Because the events are mutually disjunctive, the probability the customer is rejected upon the arrival P_{rej} (the blocking probability) can be computed as a sum of the probabilities:

$$P_{rej} = P_{full} + P_{down} = \sum_{p=0}^{n-1} P_{(m,p)} + \frac{\eta}{\eta+\zeta} \,. \tag{24}$$

3 Results of experiments

To demonstrate solvability of the model we run several testing experiments with the mathematical model. Applied values of the model parameters are summarized in Table 1.

Parameter	m [-]	λ [h^{-1}]	n [-]	$n\mu$ [h^{-1}]	η [h^{-1}]	ζ [h^{-1}]
Applied values	50	10	3	30	0.01 up to 0.1 with step 0.01	0.1 up to 1.0 with step 0.1

Table 1 The applied values of the model parameters.

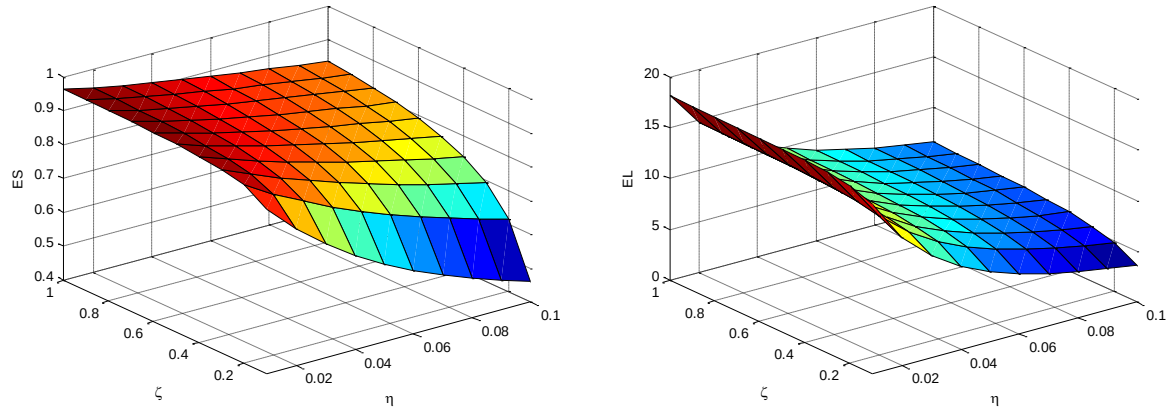


Figure 2 The dependence of *ES* and *EL* on η and ζ

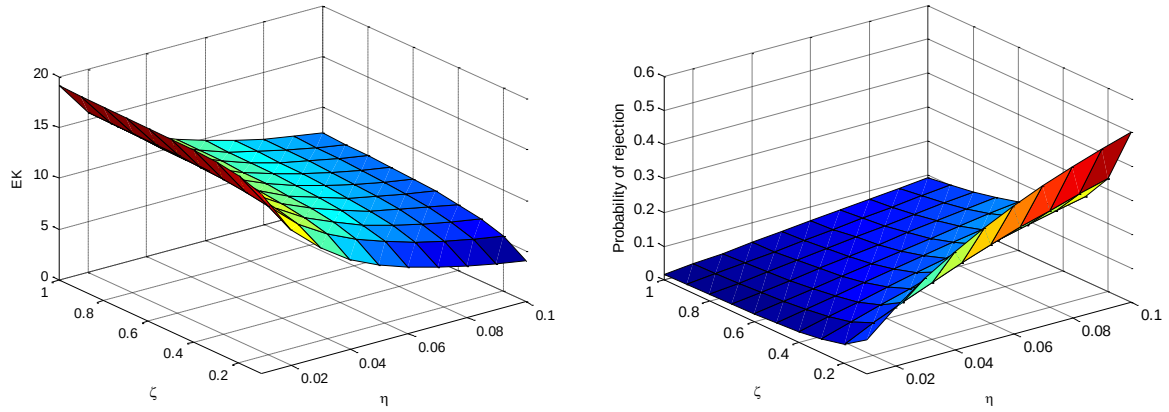


Figure 3 The dependence of *EK* and P_{rej} on η and ζ

4 Conclusions

We demonstrated the single-server queuing model subject to the catastrophes, with the input buffer with the finite capacity, and with the Erlang distributed service times. The mathematical model is solved numerically; we used Matlab to calculate the equilibrium probabilities of the individual states of the system. We demonstrated solvability of the proposed mathematical model on some numerical experiments. On the basis of them we got some dependences of the selected performance measures on the catastrophe rate and the repair rate.

In the mathematical model we assumed that the catastrophes of the server can occur when the server is either idle or busy. However, the mathematical model can be easily adapted also for different assumptions. For example, if we deleted the arc exiting the state *idle* and entering the state *down* (it is also necessary to incorporate the change into the mathematical model), we would get a queuing model subject to the catastrophes under the assumption that the catastrophes can occur only when the server is busy.

As regards our future research, we would like to apply the mathematical model in conditions of a real marshalling yard as mentioned in the introduction. In addition, we are interested also in some other performance measures such as the mean waiting time or the mean sojourn time.

References

[1] Boxma O. J., Perry D., and Stadje W.: Clearing models for M/G/1 queues. *Queuing Systems* **38** (2001), 287-306.

[2] Dorda, M., and Teichmann, D.: About a modification of $E_r/E_s/1/m$ queueing system subject to breakdowns. In: *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012, Part I.* (Ramík, J. and Stavárek, D. eds.). Silesian University in Opava, School of Business Administration in Karviná, Karviná, 2012, 117-122.

[3] Kumar B. K., Krishnamoorthy A., Madheswari S. P., and Basha S. S.: Transient analysis of a single server queue with catastrophes, failures and repairs. *Queuing Systems* **56** (2007), 133-141.

[4] Sudhesh R.: Transient analysis of a queue with system disasters and customer impatience. *Queuing Systems* **66** (2010), 95-105.

[5] Yechiali U.: Queues with system disasters and impatient customers when system is down. *Queuing Systems* **56** (2007), 195-202.

The Use of Polynomial Functions for Modelling of Mortality at the Advanced Ages

Petra Dotlačilová[1], Ondřej Šimpach[2], Jitka Langhamrová[3]

Abstract. Mortality at the advanced ages becomes more and more important topic for demographers. The extensions of the length of life is influenced mainly by the improvement in medicine. At first, there was a significant improvement in the care of infants, which caused the decrease of infant mortality. Later, it began to improve the mortality even at the advanced ages. One reason for this evolution is just a higher level of health care. Another equally important reason may be more interest in a healthy lifestyle and also a better environment. Given to this evolution, it is more and more important to have the best imagination about how long in average will live not only the youngest persons, but the oldest ones too. From the obtained results it is evident that the level of mortality of younger persons is different in comparison with the oldest ones. Therefore, it is necessary to correct the estimates of mortality at the highest ages. For this correction are used the various types of models. Given to the importance of the most accurate capture of mortality at the advanced ages are still required new models and methods that would provide the best imagination of current trends. The aim of this paper is to introduce the possibility of using of the polynomial functions for levelling and for extrapolation of mortality curves at the advanced ages. The results will be compared with the methodology, which is used by the Czech Statistical Office and with general mortality tables.

Keywords: mortality, life expectancy, polynomial functions and order, life tables

JEL Classification: C61, C63
AMS Classification: 62H12

1 Introduction

The mortality significantly affects the length of human life. Recently there occur the improvement in mortality in all age groups, but it is important to note that during the human life there become the changes in the character of mortality (Arltová, Langhamrová JI, Langhamrová JA [1] or Fiala [6]). The most different mortality trend is especially at the oldest ones (persons aged 60 and more years, see e.g. Šimpach, Dotlačilová, Langhamrová JI [14]). It is also important to realize that the data about the numbers of deaths M_x and the numbers of living (mid-year population) $\overline{S}_x$ are not exactly reliable at the highest ages (Boleslawski, Tabeau [2]) (and they can be influenced by both random and systematic error). Therefore it is important to modify mortality in some way especially at the highest ages (Gavrilov, Gavrilova [7]). There are various methods that can be used for smoothing mortality curves (e.g. model by Kannistö, Thatcher or Coale-Kisker. See e.g. Thatcher, Kanistö and Vaupel [15] or Burcin, Tesárková, Šídlo [3]). Another option is the application of polynomial functions.

We present in this paper the possibility of use of polynomial functions of different orders for the modelling of mortality curves (see e.g. Burcin, Tesárková, Šídlo [3]), and our results will be compared with the methodology of the Czech Statistical Office (see e.g. study by Koschin [9] or Šimpach [13]) and also with the general results of mortality tables. For smoothing and extrapolation of mortality curves we use a polynomial function of 2nd, 3rd and 4th order, depending on the appropriateness and significance. We use the data about the numbers of deaths $M_{x,t}$, the numbers of living $\overline{S}_{x,t}$ from the CZSO database from 1920 to 2010 for males and females in the Czech Republic, while the calculations are carried out for the 10-year time instants t = 1920, 1930, ..., 2010. Empirical, balanced and extrapolated values are presented for selected years in the synoptic charts separately for males and females. To complement our outputs there will be supplemented the values calculated according to the CZSO's methodology, which will be presented in the part called Materials and Methods. With the obtained results we lead the

[1] University of Economics in Prague, Faculty of Informatics and Statistics, Department of Mathematics, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, petra.dotlacilova@vse.cz
[2] University of Economics in Prague, Faculty of Informatics and Statistics, Dept. of Demography and Dept. of Statistics and Probability, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, ondrej.simpach@vse.cz
[3] University of Economics in Prague, Faculty of Informatics and Statistics, Department of Demography, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, langhamj@vse.cz

discussion about the suitability of a specific approach together with the recommendations applicable to the Czech population.

2 Materials and Methods

Let us denote the polynomial function of 2nd, 3rd and 4th order respectively as

$$m_{x,t} = \begin{cases} \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t \\ \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 x_t^3 + \varepsilon_t \\ \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 x_t^3 + \beta_4 x_t^4 + \varepsilon_t \end{cases}, \tag{1}$$

where $m_{x,t}$ are the age-specific mortality rates calculated by ratio

$$m_{x,t} = \frac{M_{x,t}}{\overline{S}_{x,t}}, \tag{2}$$

β_0, β_1, …, β_4 are the parameters of models and ε_t is the error term with the characteristics of white noise, x is age where $x = \langle 0 ; 110 \rangle$ and t is time where $t = \langle 1920 ; 2010 \rangle$ in 10-year time instants. These polynomial functions we use for smoothing of the age-specific mortality rates from 60 to 85 years. For the advanced ages we perform the extrapolation. The results will be subsequently compared with the mortality tables of CZSO. Their calculation can be summarized in a few steps. First, there is performed the smoothing of empirical age-specific mortality rates, followed by the calculation of mortality tables. In lower ages we smooth the age-specific mortality rates by moving averages with the length of 3 (e.g. Šimpach [13] or Šimpach, Dotlačilová, Langhamrová [14]) as

$$\tilde{m}^{(3)} = \frac{m_{x-1} + m_x + m_{x+1}}{3}, \text{ where } x \in \langle 3 ; 5 \rangle, \tag{3}$$

with the length of 9 as

$$\begin{aligned} \tilde{m}^{(9)} &= 0{,}2m_x + 0{,}16\left(m_{x-1} + m_{x+1}\right) + 0{,}12\left(m_{x.-2} + m_{x+2}\right) \\ &+ 0{,}08\left(m_{x-3} + m_{x+3}\right) + 0{,}04\left(m_{x-4} + m_{x+4}\right) \end{aligned}, \tag{4}$$

where $x \in \langle 6 ; 29 \rangle$ and with the length of 19 as

$$\begin{aligned} \tilde{m}^{(19)} &= 0{,}2m_x + 0{,}1824\left(m_{x-1} + m_{x+1}\right) + 0{,}1392\left(m_{x-2} + m_{x+2}\right) \\ &+ 0{,}0848\left(m_{x-3} + m_{x+3}\right) + 0{,}0336\left(m_{x-4} + m_{x+4}\right) - 0{,}0128\left(m_{x-6} + m_{x+6}\right), \\ &- 0{,}0144\left(m_{x-7} + m_{x+7}\right) - 0{,}0096\left(m_{x-8} + m_{x+8}\right) - 0{,}0032\left(m_{x-9} + m_{x+9}\right) \end{aligned} \tag{5}$$

where $x \in \langle 30 ; 59 \rangle$. For age 1 and 2 are used the empirical values of age-specific mortality rates. At the advanced ages is used the Gompertz–Makeham function (G–M) (Gompertz [8], Makeham [11]) or the modified Gompertz–Makeham function (mG–M) for smoothing of the age-specific mortality rates. Let us explain the G–M function (see e.g. Boleslawski, Tabeau [2]) as

$$\mu_x = a + bc^x, \tag{6}$$

where μ_x is the intensity of mortality, x is age and a, b, and c are parameters. The function is based on the assumptions that the increments of the intensity of mortality are constant with increasing age. Given that this assumption is not possible to apply from 85 years higher, therefore it was designed mG–M function. It assumes that the increments of the intensity of mortality decrease with increasing age. For the intensity of mortality at lower ages is true the equation:

$$\mu\left(x + \frac{1}{2}\right) \doteq m_x, \text{ where } x = 1, 2, \ldots, 59. \tag{7}$$

At first there will be used the Solver in MS Office Excel (version 2010 or higher, for more information see e.g. Fiala [5] Šimpach [12]). We have to find the initial estimates of parameters in the first step and we will improve them by using the ordinary least squares method (OLS). Because we need to estimate three parameters, we will need three equations. It is important to choose the beginning of the first interval x_0 (i.e. the age from which will be

performed the equalization) and set the length of the interval k. After that, we will calculate the sum of empirical mortality rates in the three intervals

$$G_1 = \sum_{x=x_0}^{x_0+k-1} m_x \,,\; G_2 = \sum_{x=x_0+k}^{x_0+2k-1} m_x \;\text{ and }\; G_3 = \sum_{x=x_0+2k}^{x_0+3k-1} m_x \,, \tag{8}$$

where $k = 5$ and $x_0 = 60$. The value of G_1 we can also express using by model parameters and age x (e.g. Fiala [5] or Šimpach [13]) as

$$G_1 = \sum_{x=x_0}^{x_0+k-1} \left(a + b.c^{\,x+\frac{1}{2}} \right). \tag{9}$$

In the same way, we could expressed G_2 and G_3. By subtracting and dividing of the individual equations we exclude the most of the parameters and we get

$$c^k = \frac{G_3 - G_2}{G_2 - G_1}. \tag{10}$$

The value of the parameter c we get as k-(th) square root of the c^k

$$c = \sqrt[k]{c^k}\,. \tag{11}$$

To estimate the initial values of a and b we have to calculate the value of sub-expression of K_c as

$$K_c = c^{\,x_0+\frac{1}{2}.\frac{c^k-1}{c-1}}, \tag{12}$$

and the parameters b and a we obtain as

$$b = \frac{G_2 - G_1}{K_c.\left(c^k - 1\right)}, \quad a = \frac{G_1 - b.K_c}{k} \quad \text{respectively.} \tag{13}$$

The last step is the calculation of the weighted squared deviations (*WSD*), through which we will optimize the parameters of G–M (respectively mG–M) model

$$WSD = \frac{S_{t,x} + S_{t+1,x}}{2.m_x.\left(1 - m_x\right)} \cdot \left(m_x - \tilde{m}_x^{(GM)} \right)^2 \quad \text{for} \quad x \in \langle 60\,;\, y \rangle, \tag{14}$$

where $S_{t,x}$ is the number of living at age x in year t, $S_{t+1,x}$ is the number of living at age x in year t+1, m_x are age-specific mortality rates, $m_x^{\sim(GM)}$ are smoothed values of age-specific mortality rates obtained by using G–M or mG–M model, y is the highest age for which we have a non-zero value of m_x. When we optimize the parameters using by the OLS, it is necessary to create two instrumental sums

$$S_1 = \sum_{x=60}^{82} \frac{S_{t,x} + S_{t+1,x}}{2.m_x.\left(1 - m_x\right)} \cdot \left(m_x - \tilde{m}_x^{(GM)} \right)^2, \tag{15}$$

respectively

$$S_2 = \sum_{x=83}^{y} \frac{S_{t,x} + S_{t+1,x}}{2.m_x.\left(1 - m_x\right)} \cdot \left(m_x - \tilde{m}_x^{(GM)} \right)^2, \tag{16}$$

which will help us to optimize the initial values of parameters. For continuation of analysis, let us explain mG–M function (see e.g. Lagerås [10] or Thatcher et al. [15]) as

$$\mu_x = a + bc^{\,x_0+\frac{1}{\gamma}\ln\left[\gamma(x-x_0)+1\right]}, \tag{17}$$

where a, b and c are the parameters of the original G–M function, x_0 is the age from which will be performed equalization and γ is a parameter of mG–M function.

3 Results and Discussion

Using the methodology of polynomial functions of various orders we smooth the mortality curves of the Czech males and females at the highest ages. This smoothing is applied to the data matrix of age-and-sex specific mortality rates in 1920, 1930, ..., 2010. Then we compare the obtained results from these functions with the results obtained by the CZSO's methodology. The aim of this paper is primarily to find out how it changes the suitability of various functions and the approaches for specific gender and time period. In the Fig. 1 is shown a matrix of charts with empirical and smoothed values of age-specific mortality rates of males in the Czech Republic. The charts below show only functions, which were based on the statistically significant parameter estimates (see Tab. 1 for males). It is clear that from the obtained results are statistically significant the polynomial functions of the 2nd and 3rd order in all analysed periods. On the other hand the polynomial function of the 4th order is significant for two years only (1960 and 1970). If we compare the various ways of smoothing, we conclude that a polynomial function is suitable for smoothing of mortality curves approximately to 80 years. On the contrary for the higher ages is better suited the polynomial function of the 3rd order. If we examine the suitability of these functions over time, we find that at the beginning of the analysed period was slightly better the polynomial function of the 2nd order, but towards presence is more suitable for our requirements the polynomial function of the 3rd order. If we compare our smoothing by polynomial functions with the results which we calculated by the Czech Statistical Office methodology, we conclude that there has been an undervaluation of actual level of mortality in certain years. Better smoothing provides the polynomial function of the 3rd order (especially at the highest ages).

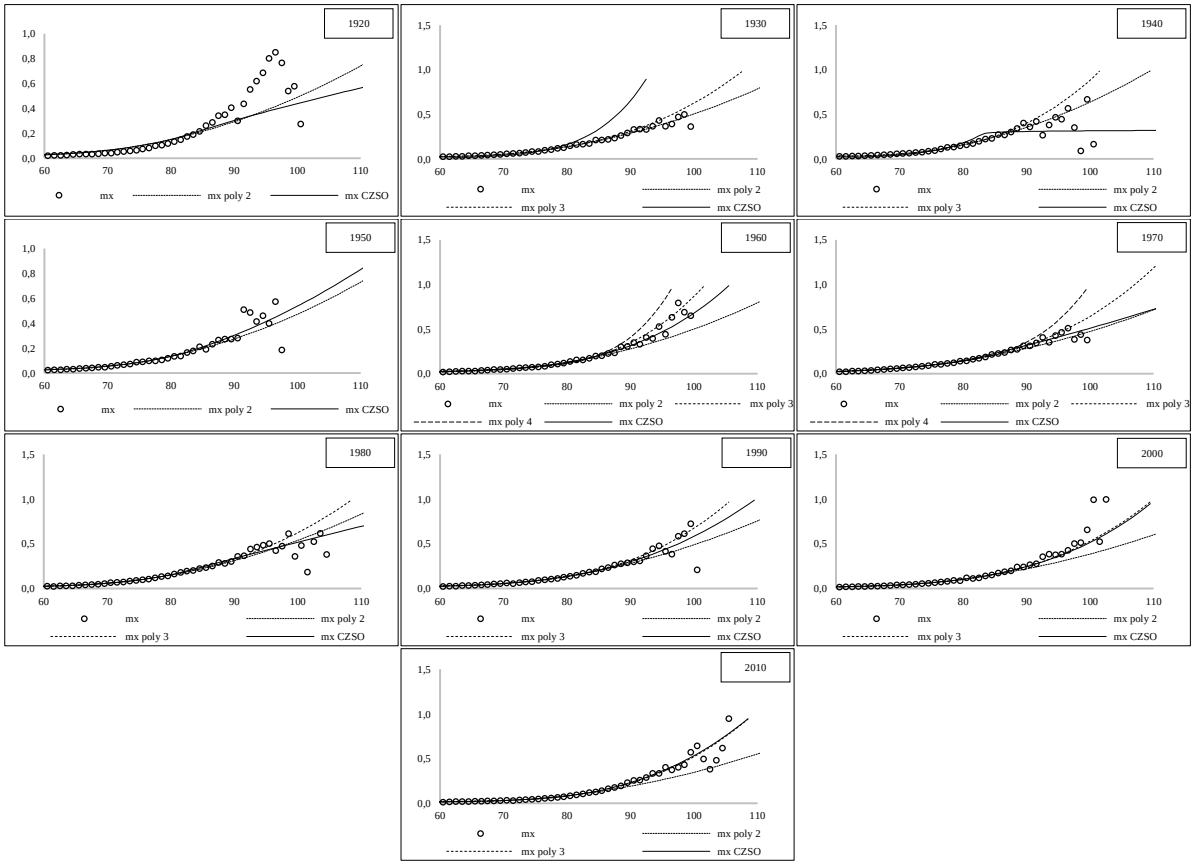


Figure 1 Age-specific mortality rates of Czech males from 1920 to 2010. Source: CZSO, own calculation

In the Fig. 2 is shown a matrix of charts with empirical and smoothed values (for parameters see Tab. 2) of age-specific mortality rates of females. We examine the various types of smoothing of these values and we come to a similar conclusion as in the case of males. The polynomial function of the 4th order is statistically significant for two years only (2000 and 2010). If we compare the smoothed values calculated by polynomial functions of 2nd and 3rd order, we find that these functions provide good estimates approximately to 85 years. From our results it is also evident that the use of polynomial functions of the 2nd and 3rd order is particularly suitable at the beginning of analysed period (approximately until 1950). Since 1960 seems to be the polynomial function of the 3rd order and CZSO's methodology as optimal smoothing approach. For smoothing of values at the advanced ages (over 85) it is more suitable the polynomial function of the 3rd order. When we compare our results calculated by the CZSO's approach, we get the simple conclusion: closer to present, the more are the values of mortality overestimated (especially the oldest ages). Then it is better to use the universal approach of polynomial function of the 3rd order.

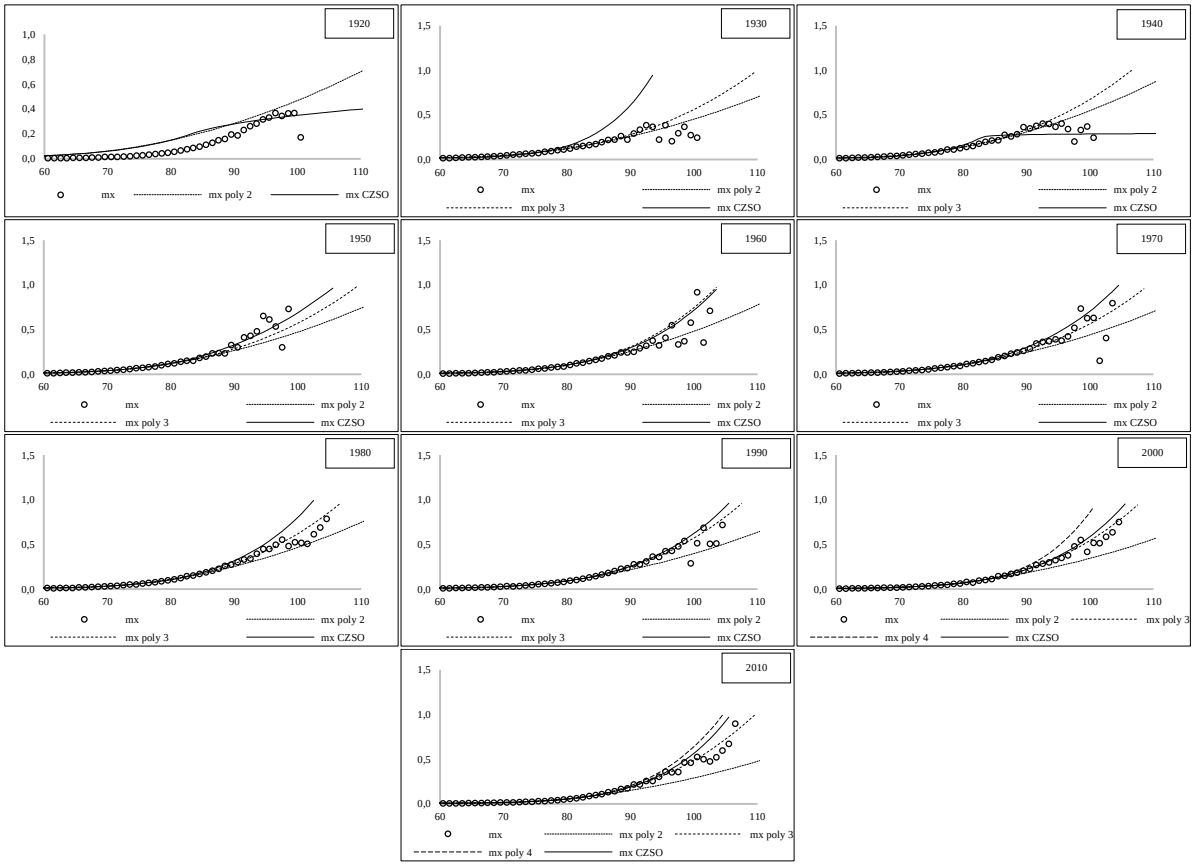


Figure 2 Age-specific mortality rates of Czech females from 1920 to 2010. Source: CZSO, own calculation

	1920	**1930**		**1940**		**1950**	**1960**			**1970**		
par.	2nd	2nd	3rd	2nd	3rd	2nd	2nd	3rd	4th	2nd	3rd	4th
β_0	9,604E-1	3,237E-4	-1,431E+0	1,774E+0	1,454E-5	1,116E+0	1,430E+0	2,155E-5	1,378E-6	9,140E-1	-2,825E+0	1,991E+1
β_1	-3,183E-2	-3,966E-2	7,253E-2	-5,556E-2	1,722E-1	-3,577E-2	-4,447E-2	-4,333E-3	-3,780E-4	-3,041E-2	1,263E-1	-1,146E+0
β_2	2,724E-4	1,246E+0	-1,234E-3	4,435E-4	-2,719E-3	2,944E-4	3,530E-4	2,931E-1	3,892E-2	2,612E-4	-1,915E-3	2,465E-2
β_3	x	x	7,161E-6	x	-3,660E+0	x	x	-6,623E+0	-1,778E+0	x	1,000E-5	-2,354E-4
β_4	x	x	x	x	x	x	x	x	3,040E+1	x	x	8,461E-7

	1980		**1990**		**2000**		**2010**	
	2nd	3rd	2nd	3rd	2nd	3rd	2nd	3rd
…	1,237E+0	-6,491E-1	1,178E+0	-2,896E+0	9,667E-1	-2,198E+0	1,019E+0	-2,837E+0
…	-4,002E-2	3,904E-2	-3,757E-2	1,332E-1	-3,073E-2	1,020E-1	-3,159E-2	1,301E-1
…	3,318E-4	-7,656E-4	3,080E-4	-2,063E-3	2,501E-4	-1,592E-3	2,496E-4	-1,994E-3
…	x	5,046E-6	x	1,090E-5	x	8,469E-6	x	1,032E-5
…	x	x	x	x	x	x	x	x

Table 1 Estimated parameters of significant polynomial functions for males. Source: author's calculations

	1920	**1930**		**1940**		**1950**		**1960**		**1970**	
par.	2nd	2nd	3rd	2nd	3rd	2nd	3rd	2nd	3rd	2nd	3rd
β_0	8,000E-1	1,059E+0	-1,341E+0	1,392E+0	-1,383E+0	1,181E+0	-1,012E+0	1,469E+0	-4,366E+0	1,176E+0	-1,529E+0
β_1	-2,744E-2	-3,421E-2	6,641E-2	-4,458E-2	7,176E-2	-3,801E-2	5,393E-2	-4,577E-2	1,988E-1	-3,750E-2	7,590E-2
β_2	2,423E-4	2,825E-4	-1,114E-3	3,628E-4	-1,252E-3	3,105E-4	-9,657E-4	3,603E-4	-3,035E-3	3,028E-4	-1,271E-3
β_3	x	x	6,422E-6	x	7,425E-6	x	5,868E-6	x	1,561E-5	x	7,238E-6
β_4	x	x	x	x	x	x	x	x	x	x	x

	1980		**1990**		**2000**			**2010**		
	2nd	3rd	2nd	3rd	2nd	3rd	4th	2nd	3rd	4th
…	1,339E+0	-2,046E+0	1,141E+0	-2,672E+0	1,093E+0	-3,306E+0	1,676E+1	9,964E-1	-3,389E+0	6,021E+0
…	-4,216E-2	9,977E-2	-3,589E-2	1,240E-1	-3,389E-2	1,505E-1	-9,718E-1	-3,050E-2	1,533E-1	-3,731E-1
…	3,356E-4	-1,635E-3	2,857E-4	-1,934E-3	2,649E-4	-2,295E-3	2,114E-2	2,351E-4	-2,317E-3	8,675E-3
…	x	9,059E-6	x	1,020E-5	x	1,177E-5	-2,047E-4	x	1,173E-5	-8,981E-5
…	x	x	x	x	x	x	7,465E-7	x	x	3,501E-7

Table 2 Estimated parameters of significant polynomial functions for females. Source: author's calculations

4 Conclusion

From the obtained results it is clear that during the analysed period occurred the changes in the suitability of the various types of functions. This is mainly due to changes in the level of mortality (which is the most relevant at the highest ages). All approaches of smoothing of age-specific mortality rates seem to be problematic at the beginning of the analysed period - mainly due to low reliability of empirical data. If we compare the results calculated by the CZSO's methodology with results of polynomial functions, we conclude that especially at the present the CZSO's methodology provides the estimates which significantly overestimate the actual mortality. On the contrary seems to be more suitable the polynomial function of the 3rd order for the majority of the analysed years. Our results also indicate that its suitability for smoothing of the age-specific mortality rates is approximately from 80 years higher. If we look at the smoothed values which we obtained by applying the polynomial function of the 2nd order, we find that its application should be up to 80 years, no more. The polynomial function of the 2nd order is not suitable for smoothing at the higher ages, because it underestimates the actual values of mortality. Based on the calculated results we can also conclude that neither one of the polynomial functions is universally applicable for smoothing in whole age range and to the subsequent extrapolation of mortality curves to the highest ages. On the contrary it might be appropriate to use a combination of both types of functions for smoothing of mortality curves. Up to the age of 80 years we could applied the polynomial function of the 2nd order and at the higher ages the polynomial function of the 3rd order. This idea is also a challenge for our future research.

Acknowledgements

This paper was supported by the project of Internal Grant Agency of VŠE Praha under the No. IGS F4/24/2013.

References

[1] Arltová, M., Langhamrová, JI, Langhamrová, JA. Development of life expectancy in the Czech Republic in years 1920-2010 with an outlook to 2050. *Prague Economic Papers*, 2013, Vol. 22, no. 1, pp. 125–143.

[2] Boleslawski, L., Tabeau, E. Comparing Theoretical Age Patterns of Mortality Beyond the Age of 80. In: *Tabeau, E., Van Den Berg Jeths, A. and Heathcote, CH. (Eds.). Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*, 2001, pp. 127 – 155.

[3] Burcin, B., Tesárková, K., Šídlo, L. Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. *Demografie* **52**, 2010, pp. 77 – 89.

[4] Ekonomov, AL., Yarigin, VN. The age dynamics of mortality and the Gompertz-Makeham law, *Zhurnal Obshchei Biologii*, Vol. 50, Issue 2, MAR-APR 1989, pp. 236–243.

[5] Fiala, T. Mortality in the Age from 20 to 70 Years in the Czech Republic in the Period 1950 to 1996. *Acta Oeconomica Pragensia*, 1999, Vol. 7, no. 4, pp. 35–44.

[6] Fiala, T. Reducing the Consequences of Population Ageing for the Pension System by Raising the Retirement Age in the Czech Republic. In: *The 7th International Days of Statistics and Economics*. Slaný: MELANDRIUM, 2013, pp. 398–407.

[7] Gavrilov, L.A., Gavrilova, N.S. Mortality measurement at advanced ages: a study of social security administration death master file. *North American actuarial journal*, vol. 15, no. 3, pp. 432 – 447.

[8] Gompertz, B. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London* **115** (1825), pp. 513–585.

[9] Koschin, F. Jak vysoká je intenzita úmrtnosti na konci lidského života? *Demografie*, 1999, Vol. 41, no. 2, pp. 105–109.

[10] Lagerås, A.N. Commutation functions under Gompertz–Makeham mortality, *Scandinavian Actuarial Journal*, Vol. 2010, Issue 2, pp. 161–164.

[11] Makeham, W.M. On the Law of Mortality and the Construction of Annuity Tables. *The Assurance Magazine, and Journal of the Institute of Actuaries* **8** (1860), pp. 301–310.

[12] Šimpach, O. Faster convergence for estimates of parameters of Gompertz-Makeham function using available methods in solver MS Excel 2010. In: *Proceedings of the 30th International Conference of Mathematical Methods in Economics*. Opava: Universita Opava, 2012, pp. 870–874.

[13] Šimpach, O. Effective use of MS Excel in education of economic demographers. In: *Efficiency and Responsibility in Education*. Praha: Czech University of Life Sciences in Prague, 2013, pp. 557–564.

[14] Šimpach, O., Dotlačilová, P., Langhamrová, JI. Logistic and ARIMA models in the Estimation of Life Expectancy in the Czech Republic. In: *Proceedings of the 31st International Conference of Mathematical Methods in Economics*. Jihlava: College of Polytechnics Jihlava, 2013, pp. 915–920.

[15] Thatcher, R.A., Kanistö, V. and Vaupel, J.W. *The Force of Mortality at Ages 80 to 120*. Odense University Press, 1998.

On the Crossing Number of Cartesian Products

Emília Draženská[1]

Abstract.

The crossing numbers of Cartesian products of cycles with all graphs of order at most four are given. There are known the crossing numbers of Cartesian products of several graphs on five or on six vertices with cycles. We extend these results by giving the exact values or upper bounds of crossing numbers of Cartesian products of trees on seven vertices with cycles.

Keywords: graph, drawing, crossing number, Cartesian product, cycles.

JEL classification: C02
AMS classification: 05C10; 05C38

1 Introduction

We consider finite, undirected, simple *graph* G whose vertex set is denoted by V and edge set E. The graph is ordered pair $G = (V, E)$. A *drawing* D of the graph G in the plane is one-to-one mapping of a graph in such a way, that each vertex maps into a point in $\mathbb{R}^2$ and each edge maps into a curve between its two endpoints. The *crossing number* $cr(G)$ of a graph G is the minimum number of edge crossings in any drawing of G in the plane. The drawing with a minimum number of crossings, must be a *good drawing*, it means, that a drawing satisfied the following conditions: (i) no edge crosses itself, (ii) adjacent edges do not cross, and (iii) no two edges cross more than once.

The search for formulas for minimal number of crossings was initated by Hungarian mathematician Pal Turan [23]. In 1940 he was sent to labor camp outside Budapest. He worked in a brick factory. There were some kilns where the bricks were made and some open storage yards where the bricks were stored. All the kilns were connected to all the storage yards. The bricks were carried on small wheeled trucks to the storage yards. He had to put the bricks on the trucks at the kilns, push the trucks to the storage yards, and unload them there. The work was not difficult but the problem was at the crossings. The trucks jumped the rails, and the bricks fell out from them. He was ask: what is the minimal number of crossings? Turan realized that the actual situation could have been improved, but the exact solution of the general problem with m kilns and n storage yards seemed to be very difficult.

In graph theory, we represented the kilns and storage yards by vertices and the tracks by edges and we are asking what is the minimum number of crossings of the complete bipartite graph $K_{m,n}$ which has two sets of vertices, one with m vertices and the other one with n vertices such, that each vertex in one set is joined to every vertex in the other set.

Turan devised a drawing of $K_{m,n}$ with $\lfloor \frac{m}{2} \rfloor \lfloor \frac{m-1}{2} \rfloor \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-2}{2} \rfloor$ crossings, but the conjecture of Zarankiewicz [25] that such a drawing is the best possible, is still opened. The crossing numberof $K_{m,n}$ is proved for every n and for $1 \leq m \leq 6$ [13]. In [24] Woodall publised that the crossing number of $K_{m,n}$ is equal to Zarankiewicz number for $7 \leq m \leq 8$ and $7 \leq n \leq 10$.

The crossing numbers has been studied to improve the readability of hierarchical structures and by computer scientists and VLSI communities. Many researchers have focused on optimizing the VSLI circuit layout. VLSI theory has expanded in many directions. One of them is layout theory which studied the efficiency of embedding the graphs in the plane according to VLSI rules. One of the major tasks is minimizing the number of wire crossings in a circuit. C. Thompson introduced a graph-theoretic model for very large integration circuitry [22].

In general, compute the crossing number for a given graph is a very difficult problem. Garey and Johnson [9] have proved that this problem is NP–complete.

[1]Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovak Republic, e-mail: emilia.drazenska@tuke.sk

The crossing numbers of a few families of graphs are known. Most of them are Cartesian products of two graphs. The *Cartesian product* $G_1 \Box G_2$ of two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is a graph with vertex set $V = V_1 \Box V_2$ and edge set $E = \{\{(x_1, y_1), (x_2, y_2)\}; x_1 = x_2$ and $\{y_1, y_2\} \in E_2$ or $y_1 = y_2$ and $\{x_1, x_2\} \in E_1\}$.

Let C_n be the cycle of length n and S_n be the star isomorphic to complete bipartite graph $K_{1,n}$. Harary et al. [11] conjectured that the crossing number of $C_m \Box C_n$ is $(m-2)n$ for all m, n satisfying $3 \leq m \leq n$. This has been proved only for m, n satisfying $n \geq m,\ m \leq 7$, [1], [2], [3], [4], [5], [19], [20], [21]. It was recently proved by Glebsky and Salazar [10] that the crossing number of $C_m \Box C_n$ equals its long-conjectured value at least for $n \geq m(m+1)$. Beineke and Ringeisen in [4], Jendrol' and Ščerbová in [12] determined the crossing numbers of the Cartesian products of all graphs on four vertices with cycles. Klešč in [14], [15], [16], [17], Klešč, Richter and Stobert in [19], and Klešč and Kocúrová in [18] gave the crossing numbers of $G \Box C_n$ for several graphs of order five.

In [6], [7] and [8] there are given the crossing numbers of $G \Box C_n$ for several graph G on six vertices. It seems natural to enquire about crossing numbers of Cartesian products for other graphs with cycles. We extend these results by giving the exact values or upper bounds of crossing numbers of Cartesian products $G \Box C_n$ for every tree G on seven vertices. There are 11 trees of order seven (see Figure 1).

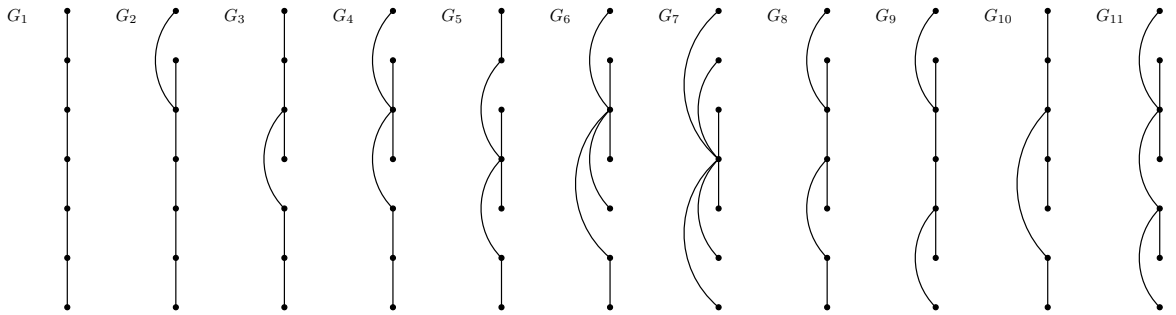


Figure 1

2 The crossing number of the Cartesian products of seven vertex trees with cycles

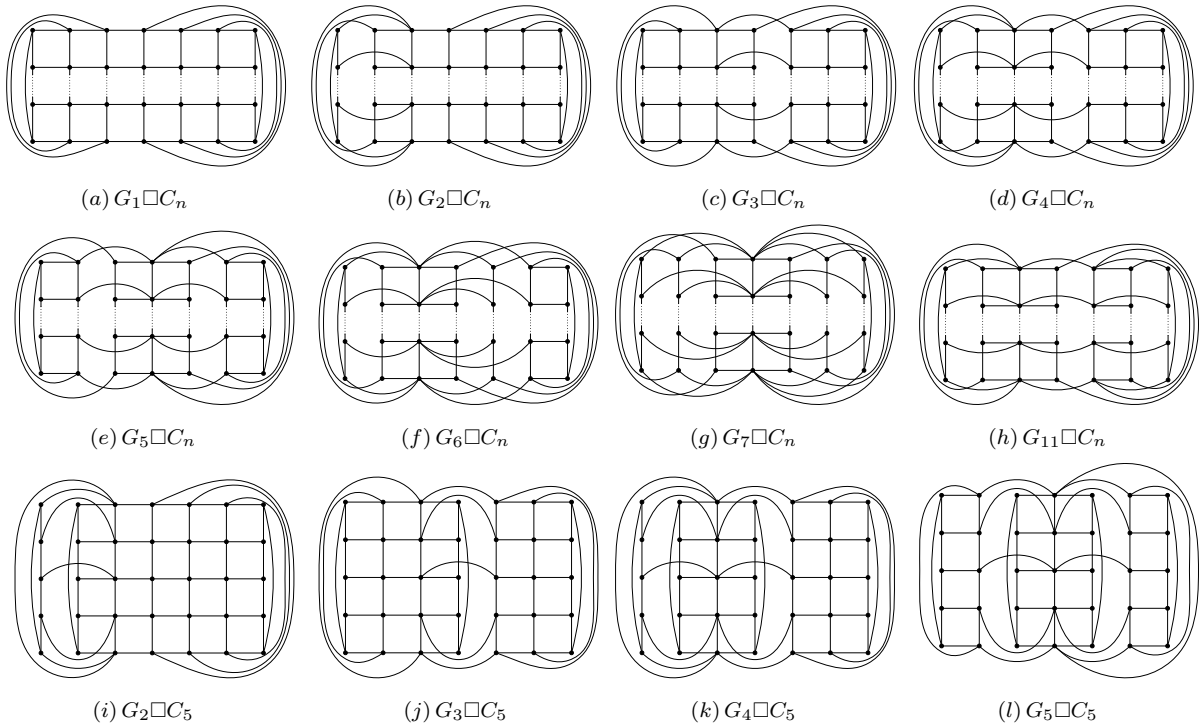
Theorem 1. *$cr(G_2 \Box C_3) = cr(G_3 \Box C_3) = 1$, $cr(G_2 \Box C_4) = cr(G_3 \Box C_4) = 2$, $cr(G_2 \Box C_5) = cr(G_3 \Box C_5) = 4$, $cr(G_4 \Box C_3) = cr(G_5 \Box C_3) = 2$, $cr(G_4 \Box C_4) = cr(G_5 \Box C_4) = 4$, $cr(G_4 \Box C_5) = cr(G_5 \Box C_5) = 8$. For $n \geq 6$, $cr(G_2 \Box C_n) = cr(G_3 \Box C_n) = n$, $cr(G_4 \Box C_n) = cr(G_5 \Box C_n) = 2n$. For $n \geq 3$, $cr(G_1 \Box C_n) = 0$, $cr(G_6 \Box C_n) \leq 4n$, $cr(G_7 \Box C_n) \leq 6n$, and $cr(G_{11} \Box C_n) \leq 3n$.*

Proof: Every graph $G_j \Box C_n$ contains n copies of the graph G_j. If we would like to find an upper bound for the crossing number for the graph $G_j \Box C_n$, it is enough to find a drawing of $G_j \Box C_n$ and a number of crossings in this drawing is an uppper bound. On the other hand, if graph $G_j \Box C_n$ contains some graph as its subgraph, for which crossing number is published, then we have a lower bound for the crossing number for the graph $G_j \Box C_n$. There is (see Figure 2(a)) a drawing of the graph $G_1 \Box C_n$ without crossings, so it is a planar graph and hence $cr(G_1 \Box C_n) = 0$.

Figures 2(i), 2(j), 2(k) and 2(l) show the drawings of the graphs $G_2 \Box C_5$, $G_3 \Box C_5$, $G_4 \Box C_5$ and $G_5 \Box C_5$ with 4, 4, 8 and 8 crossings. By deleting the edges of a middle copies of the subgraphs isomorphic to G_2, G_3, G_4 and G_5 in these drawings, we obtain the drawings of the graphs $G_2 \Box C_4$, $G_3 \Box C_4$, $G_4 \Box C_4$ and $G_5 \Box C_4$ with 2, 2, 4 and 4 crossings. And in next step, by deleting the edges of second one copies of the subgraphs isomorphic to G_2, G_3, G_4 and G_5 in these drawings, we obtain the drawing of the graphs $G_2 \Box C_3$, $G_3 \Box C_3$, $G_4 \Box C_3$ and $G_5 \Box C_3$ with 1, 1, 2 and 2 crossings. So, we have the upper bounds for the crossing numbers of graphs $G_2 \Box C_n$, $G_3 \Box C_n$, $G_4 \Box C_n$ and $G_5 \Box C_n$ for $3 \leq n \leq 5$.

Figure 2(b) shows the drawing of the graph $G_2 \Box C_n$ in which the edges of n copies of subgraph isomorphic to G_2 are crossed once, so the crossing number of $G_2 \Box C_n$ is not greater than n. Using the same reason and Figure 2(c), the crossing number of $G_3 \Box C_n$ is at most n. Figures 2(d) and 2(e) show the drawings of the graphs $G_4 \Box C_n$ and $G_5 \Box C_n$ in which the edges of n copies of subgraphs isomorphic

to G_4 and G_5 are crossed twice, so $cr(G_4\Box C_n) \leq 2n$ and $cr(G_5\Box C_n) \leq 2n$. The drawings in Figures 2(f), 2(g) and 2(h) show the graphs $G_6\Box C_n$, $G_7\Box C_n$ and $G_{11}\Box C_n$ with $4n$, $6n$ and $3n$ crossings. So, $cr(G_6\Box C_n) \leq 4n$, $cr(G_7\Box C_n) \leq 6n$ and $cr(G_{11}\Box C_n) \leq 3n$.



(a) $G_1\Box C_n$ (b) $G_2\Box C_n$ (c) $G_3\Box C_n$ (d) $G_4\Box C_n$

(e) $G_5\Box C_n$ (f) $G_6\Box C_n$ (g) $G_7\Box C_n$ (h) $G_{11}\Box C_n$

(i) $G_2\Box C_5$ (j) $G_3\Box C_5$ (k) $G_4\Box C_5$ (l) $G_5\Box C_5$

Figure 2

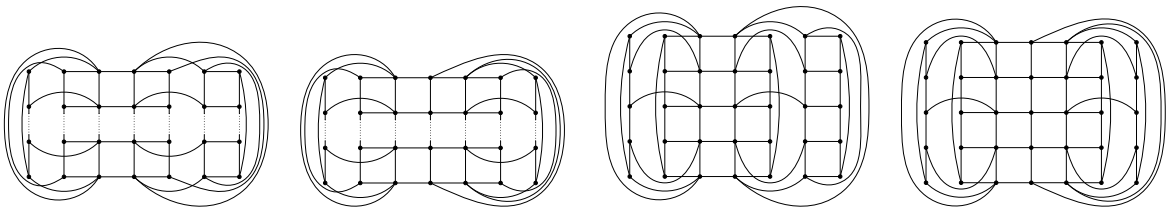
Next, we find the lower bounds of crossing numbers $cr(G_j\Box C_n)$ for $j = 2, 3, 4, 5$. The graph $K_{1,3}\Box C_n$ is a subgraph of the graph $G_2\Box C_n$ and also of the graph $G_3\Box C_n$ and in [12] it is proved that $cr(K_{1,3}\Box C_3) = 1$, $cr(K_{1,3}\Box C_4) = 2$, $cr(K_{1,3}\Box C_5) = 4$ and $cr(K_{1,3}\Box C_n) = n$ for $n \geq 6$. We have the lower bounds 1, 2, 4 and n for the crossing numbers of $G_2\Box C_n$ and of the graph $G_3\Box C_n$. The graph $K_{1,4}\Box C_n$ is a subgraph of the graphs $G_4\Box C_n$ and $G_5\Box C_n$, and Klešč in [14] published that $cr(K_{1,4}\Box C_3) = 2$, $cr(K_{1,4}\Box C_4) = 4$, $cr(K_{1,4}\Box C_5) = 8$ and $cr(K_{1,3}\Box C_n) = 2n$ for $n \geq 6$. So, we have the lower bounds 2, 4, 8 and $2n$ for the crossing numbers of $G_4\Box C_n$ and of the graph $G_5\Box C_n$. The upper and lower bounds of crossing numbers for graphs $G_j\Box C_n$ for $j = 2, 3, 4, 5$ are the same. So, we have values of crossing numbers of corresponding graphs. □

Theorem 2. *$cr(G_8\Box C_3) = cr(G_9\Box C_3) = 2$, $cr(G_8\Box C_4) = cr(G_9\Box C_4) = 4$, $cr(G_8\Box C_5) = cr(G_9\Box C_5) = 8$, $cr(G_8\Box C_n) = cr(G_9\Box C_n) = 2n$ for $n \geq 6$.*

Proof: The drawing of the graph $G_8\Box C_n$ in Figure 3(a) shows that the edges of n copies of subgraph isomorphic to G_8 are crossed twice, hence the crossing number of $G_8\Box C_n$ is at most $2n$. Figure 3(b) shows the drawing of the graph $G_9\Box C_n$ with $2n$ crossings, so $cr(G_9\Box C_n) \leq 2n$. Using Figure 3(c), we have $cr(G_9\Box C_5) \leq 8$ and the removing the edges of subgraph isomorphic to G_8 in the middle results in the drawing of a subdivision of the graph $G_8\Box C_4$ with four crossings. Subsenquently, the removing the edges of subgraph isomorphic to G_8 down, we have drawing of a subdivision of the graph $G_8\Box C_3$ with two crossing. Similar methods in Figure 3(d) where is drawing of graph $G_9\Box C_5$, we have $cr(G_9\Box C_5) \leq 8$, $cr(G_9\Box C_4) \leq 4$ and $cr(G_9\Box C_3) \leq 2$.

By the removal of vertex of degree one which is adjacent with vertex of degree two in the graph G_8, we obtain a graph, denoted it by H. In [8] it is proved, that $cr(H\Box C_3) = 2$, $cr(H\Box C_4) = 4$, $cr(H\Box C_5) = 8$ and $cr(H\Box C_n) = 2n$ for $n \geq 6$, so we have lower bounds for the crossing numbers of $G_8\Box C_n$ for every $n \geq 3$. Graph G_9 is a subdivision of the graph H, hence $cr(G_9\Box C_3) \geq 2$, $cr(G_9\Box C_4) \geq 4$, $cr(G_9\Box C_5) \geq 8$

and $cr(G_9 \square C_n) \geq 2n$ for $n \geq 6$. □



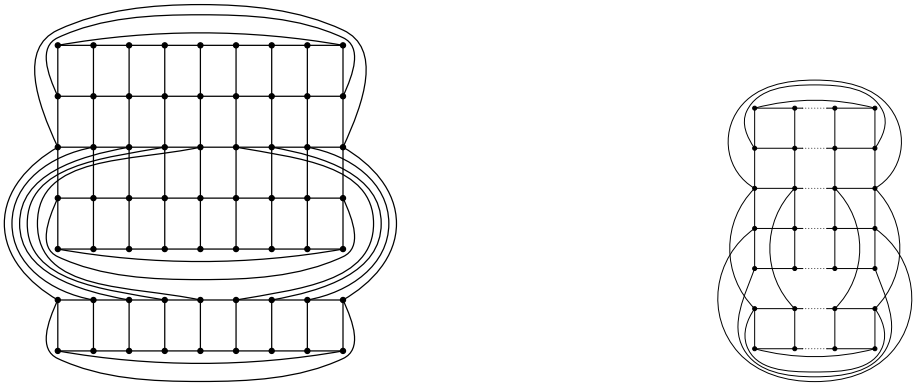
(a) $G_8 \square C_n$ (b) $G_9 \square C_n$ (c) $G_8 \square C_5$ (d) $G_9 \square C_5$

Figure 3

Theorem 3. $cr(G_{10} \square C_3) = 1$, $cr(G_{10} \square C_4) = 2$, $cr(G_{10} \square C_5) = 4$, $cr(G_{10} \square C_6) = 6$, $cr(G_{10} \square C_7) \leq 9$, $cr(G_{10} \square C_8) \leq 12$, $cr(G_{10} \square C_9) \leq 16$, $cr(G_{10} \square C_n) \leq 2n$, *for* $n \geq 10$.

Proof: In Figure 4(a) there is a drawing of the graph $G_{10} \square C_9$ with 16 crossings. By deleting the edges of copies of subgraph isomorphic to G_{10} alternately from the left and from the right, we obtain drawings of the graphs $G_{10} \square C_i$ for $i = 8, 7, 6, 5, 4, 3$ with 12, 9, 6, 4, 2 and 1 crossings. The drawing of the graph $G_{10} \square C_n$ with $2n$ crossings, in which every copy of subgraph isomorphic to G_{10} is crosses two times, is in Figure 4(b).

Removing of vertex of degree one in the graph G_{10}, we obtain a graph, denoted it by F. In [8] it is proved, that $cr(F \square C_3) = 1$, $cr(F \square C_4) = 2$, $cr(F \square C_5) = 4$ and $cr(F \square C_n) = n$ for $n \geq 6$. We have lower bounds for the crossing numbers of $G_{10} \square C_n$ for every $n \geq 3$. Then we obtain the crossing numbers of graphs $G_{10} \square C_n$ for $n = 3, 4, 5, 6$ are respectively 1,2, 4 and 6. □



(a) $G_{10} \square C_7$ (b) $G_{10} \square C_n$

Figure 4

References

[1] Adamson, J., Richter, R. B.: Arrangements, circular arrangements and the crossing number of $C_7 \times C_n$, *J. Combinatorial Theory-B* **90** (2004), 21–39.

[2] Anderson, M., Richter, R. B., Rodnay, P.: The crossing number of $C_6 \times C_6$, *Congr. Numerantium* **118** (1996), 97–107.

[3] Anderson, M., Richter, R. B., Rodnay, P.: The crossing number of $C_7 \times C_7$, *Congr. Numerantium* **125** (1997), 97–117.

[4] Beineke, L. W., RINGEISEN, R. D.: On the crossing numbers of products of cycles and graphs of order four, *J. Graph Theory* **4** (1980), 145–155.

[5] Dean, A. M., Richter, R. B.: The crossing number of $C_4 \times C_4$, *J. Graph Theory* **19** (1995), 125–129.

[6] Draženská, E.: The crossing number of $G\Box C_n$ for the graph G on six vertices, *Mathematica Slovaca*, **61** (2011), 675–686.

[7] Draženská, E., Klešč, M.: On the crossing numbers of $G\Box C_n$ for graphs G on six vertices, *Discussiones Mathematicae Graph Theory*,**36** (2007), 109–119.

[8] Draženská, E., Klešč, M.: The crossing numbers of products of cycles with 6–vertex trees, *Tatra Mountains Math. Publ.*, **36** (2007), 109–119.

[9] Garey, M. R., Johnson, D. S.: Crossing number is NP-complete, *SIAM J. Algebraic and Discrete Methods* **4**, (1983) 312–316.

[10] Glebsky, L. Y., Salazar, G.: The crossing number of $C_m \times C_n$ is as conjectured for $n \geq m(m+1)$, *J. Graph Theory* **47** (2004), 53–72.

[11] Harary, F., Kainen, P. C., Schwenk, A. J.: Toroidal graphs with arbitrarily high crossing numbers, *Nanta Math.* **6** (1973), 58–67.

[12] Jendroľ, S., Ščerbová,M.: On the crossing numbers of $S_m \times P_n$ and $S_m \times C_n$, *Časopis pro pěstování matematiky* **107** (1982), 225–230.

[13] Kleitman, D. J.: The crossing number of $K_{5,n}$ *J. Comb. Theory* **9**, (1970) 315–323.

[14] Klešč, M.: On the crossing numbers of Cartesian products of stars and paths or cycles, *Mathematica Slovaca* **41** (1991), 113–120.

[15] Klešč, M.: The crossing numbers of Cartesian products of paths with 5–vertex graphs, *Discrete Mathematics* **233** (2001), 353–359.

[16] Klešč, M.: The crossing number of $K_{2,3} \times C_3$, *Discrete Mathematics* **251** (2002), 109–117.

[17] Klešč, M.: Some crossing numbers of products of cycles, *Discussiones Mathematicae - Graph theory* **25** (2005), 197–210.

[18] Klešč, M., Kocúrová, A.:The crossing number of products of 5-vertex graphs with cycles, *Discrete Mathematics* **307** (2007), 1395–1403.

[19] Klešč, M., Richter, R. B., Stolbert, I.: The crossing number of $C_5 \times C_n$, *J. Graph Theory* **22** (1996), 239–243.

[20] Richter, R. B., Salazar, G: The crossing number of $C_6 \times C_n$, *Australasian Journal of Combinatorics* **23** (2001), 135–144.

[21] Ringeisen, R. D., Beineke, L. W.: The crossing number of $C_3 \times C_n$, *J. Combinatorial Theory* **24** (1978), 134–136.

[22] Thompson, C. D.: "Area-time complexity for VLSI", *Proceedings of the eleventh annual ACM symposium on Theory of computing, ser. STOC 79*, (1979) 81–88.

[23] Turán, P.: "A note of welcome", *J. Graph Theory* **1**, (1977) 7–9.

[24] Woodall, D.R.: Cyclic–order graphs and Zarankiewicz's crossing–number conjecture, *J. Graph Theory* **17** (1993), 657–671.

[25] Zarankiewicz, K.: "On a problem of P. Turn concerning graph", *Fund. Math.* **41**, (1977) 137–145.

Multifactor dynamic credit risk model

Jaroslav Dufek[1], Martin Šmíd[2]

Abstract.

We propose a new dynamic model of the Merton type, based on the Vasicek model. We generalize Vasicek model in three ways: we add model for loss given default (LGD), we add dynamics to the model and we allow non-normal distributions of risk factors. Then we add a retrospective interaction of underlying factors and found a non-linear behaviour of these factors.

In particular, the evolution of factors underlying the DR and the LGD is assumed to be ruled by a non-linear vector AR process with lagged DR and LGD and their non-linear transformations.

We apply our new model on real US mortgage data and demonstrate its statistical significance.

Keywords: loss given default, default rate, credit risk.

JEL classification: G21, C58
AMS classification: 62P20

1 Introduction

Asking a question "Why do we do this?" the answer could be: because the risk which follows from the real estate market is bigger than what was expected. This was shown in a recent crisis a few years ago.

Our study is based on the famous Vasicek model. Modifications of the Vasicek model are in plentiful supply (for example [2], [5] or [1] etc.). In [1] is quite huge literature review for more details see there. But on the other hand just a few of these modifications are dynamic ones (for example [3]). So we will focus on a rising new dynamic model of the Merton type, based on the Vasicek model.

There is a vast amount of literature in this area of interest. By far the most famous and most frequently used model is the Vasicek model for default rate, see [6]. The Vasicek assuming a fixed LGD. There are many models for random LDG, see [5], [1], [2], [3] and references therein. The model in [3] is a dynamic one based on the Vasicek model.

Now we will mention several studies which react on recent crisis. In [7], [9] and [4] is shown that LGD non-linearly depends on house price index and its history, which is not surprising. In [8] is summary of current state of art at the mortgage risk modelling.

Our model is based on [3]. We use the same structure: models for the default rate and for the LGD are same. Our original contribution is the creation of sub-models for underlying factors. We want to model the situation when bank losses retrospectively affect the default rate. Then we obtain a non-linear model.

2 Definition of the model

We want to model a situation when we have one creditor (for example a bank) with a countable number n of debtors (clients). The value of the i-th debtor's assets at time t is $A_{i,t}$. We assume that each debtor pays a regular instalment b.

[1]Department of Probability and Mathematical Statistics, MFF UK, Prague, The Czech Republic, jaroslav.dufek.2@seznam.cz
[2]Institute of Information Theory and Automation, Academy of Sciences of The Czech Republic, martinsmid.eu@gmail.com

The default of the i-th debtor is a state when the value of assets decrease under a given threshold B_i. Then the definition of the probability of default at time t is $P[A_{i,t} < B_i]$. We will use DR_t – the default rate – very frequently. The default rate is a simple ratio

$$DR_t = \frac{number\ of\ defaults}{number\ of\ loans} \tag{1}$$

2.1 Model for DR – default rate

We assume that

$$\log A_{i,t} = \log A_{i,t-1} + \Delta Y_t + \Delta V_{i,t}, \qquad i \leq n \tag{2}$$

where n is a number of debtors, $A_{i,t}$ is a value of assets of i-th debtor at time t, and $\Delta Y_t := Y_t - Y_{t-1}$, Y_t is a common factor following the stochastic process.

We assume that the duration of the debt is just one period and that the value of assets in each period is

$$\log A_{i,t-1} = Y_{t-1} + V_{i,t-1} \qquad i \leq n, \tag{3}$$

where $V_{i,t}$ is r. v. specific to the i-th debtor. We assume that $\{V_{i,t}\}_{i \leq n,\ t \in \mathbb{N}}$ are mutually independent and independent with respect to $\Delta Y_t,\ t \in \mathbb{N}$.

From (2) and (3), and from the assumption of independence, we can obtain that conditional probability of the default of the i-th debtor at time t for a given $\overline{Y}_t := (\Delta Y_1, \dots, \Delta Y_{t-1})$ is

$$P[A_{i,t} < B_i | \overline{Y_t}] = P[\Delta V_{i,t} + V_{i,t-1} < \log B_i - Y_t | \overline{Y_t}] = \Psi(\log B_i - Y_t), \tag{4}$$

where Ψ is the distribution function of r. v. $V_{i,t}$ which is identically distributed with $\mathbb{E}V_{i,t} = 0$ and $\mathrm{var}V_{i,t} = \sigma^2,\ \sigma > 0$.

If we assume that the debts are identical for all periods - $\log B_{i,t} = b$, if we approximate $DR_t = \lim_n \frac{\text{number of defaults at time } t}{n}$ we may apply the Law of Large Numbers to the conditional probabilities from (3) (we can do this when $A_{1,t},\ A_{2,t}, \dots$ are conditionally independent with respect to $\overline{Y_t}$), then we obtain $DR_t = P[A_{i,t} < b | \overline{Y_t}] \doteq \Psi(b - Y_t)\ t \in \mathbb{N}$, further implying that (under assumption that function Ψ is monotonic)

$$\Delta Y_t \doteq \Psi^{-1}(DR_{t-1}) - \Psi^{-1}(DR_t). \tag{5}$$

Let us note that Ψ is a general distribution function but for our valuation we will assume that $\Psi(x) = \Phi(x)$, where Φ is a distribution function of a standard normal distribution.

2.2 Model for loss

Now we will introduce our model for loss of the bank. From formula (14) in [3] we have that

$$L_t = DR_t \cdot h(I_t), \tag{6}$$

where L_t is the realised bank loss, DR_t is the default rate and I_t represents the price index of properties. From formula (17) in [3] we directly obtain

$$h(t) = \Phi(\frac{-t}{\sigma}) - \exp\{t + \frac{1}{2}\sigma^2\}\Phi(\frac{-t}{\sigma} - \sigma). \tag{7}$$

Justification (under the assumption that a property price follows geometric Brownian motion and $h(t) = 1 - RR(t)$, where RR is a recovery rate) and valuation of the function h is in Appendix in [3].

2.3 Analysis of function h

The function h is one of the main pillars of our model; so the behaviour of this function is very important to us. The function h is a convex-concave function, so its inflexion point is the most important to us

because changes in house price index has the biggest impact to loss in neighbourhood of its inflexion point. After some algebra we obtain the first and second derivations:

$$h(t) = \Phi(\frac{-t}{\sigma}) - \exp\{t + \frac{1}{2}\sigma^2\}\Phi(\frac{-t}{\sigma} - \sigma), \tag{8}$$

$$h'(t) = -\exp t + \frac{1}{2}\sigma^2\Phi(\frac{-t}{\sigma} - \sigma), \tag{9}$$

$$h''(t) = \frac{-1}{\sigma\sqrt{2\pi}}\exp\frac{t^2}{2\sigma^2} - \exp t + \frac{1}{2}\sigma\Phi(\frac{-t}{\sigma} - \sigma). \tag{10}$$

But we can't obtain the inflexion point analytically because of the Equation (11), which is equivalent to $h''(t) = 0$.

$$\Phi(\frac{-t}{\sigma} - \sigma) = \frac{1}{\sigma}\varphi(\frac{-t}{\sigma} - \sigma). \tag{11}$$

We can obtain it only numerically. We know that the inflexion point is unique from the graphical solution.

2.4 Evolution of factors

When we consider about evolution of underlining factors we try to model situation that default rate depend on house price index, loss of the bank and theirs previous values and itself lagged values. Thus we assume that the number of people who are not able to pay their loan is growing significantly, the ratio of unpaid loans increases in all banks. Banks have their investments covered by real properties so they are losing part of their liquidity. If a bank wants to recover lost liquidity it must sell some of its real properties, if all banks chose this strategy, the value would decrease, equity would not be sufficient and the LGD would increase.

We assume that common factors Y_t and I_t are driven by these equations:

$$\Delta Y_t = C_1 + a_1\Delta Y_{t-1} + b_1\Delta Y_{t-2} + \underbrace{c_1\Delta L_{t-3} + d_1\Delta L_{t-4}}_{\text{retrospective interaction}} + \\ + e_1\Delta I_{t-2} + \varepsilon_{1,t} \tag{12}$$

$$\Delta I_t = C_2 + a_2\Delta Y_{t-2} + b_2\Delta Y_{t-3} + c_2\Delta DR_{t-3} + d_2\Delta DR_{t-4} + \\ + e_2\Delta I_{t-1} + f_2 I_{t-2} + g_2\Delta I_{t-3} + \varepsilon_{2,t} \tag{13}$$

where ε_t are iid independent, non-correlated and normally distributed.

3 Empirical results

We tested our proposed model on a real dataset which is described below.

3.1 Description of the data set

The dataset for our empirical work contains quarterly delinquency rates[1] on mortgage loans from the US economy, which are provided by the US Department of Housing and Urban Development and the Mortgage Bankers Association.[2] We used the Standard & Poor Price Index of properties. The data for the default rate starts in the first quarter of 1979 and ends in the first quarter of 2012. The data for house price index starts in the first quarter 1987 and ends in the third quarter of 2012. We will use our data only from the first quarter 1987 forward (due to missing values for the house price index prior to that date).

In Figure 1 we can see a peak in 2008 which corresponds to the recent crisis in 2009. That is quite interesting, because in Figure 2 we can see a peak in 2010 (values of foreclosures are in percent); so the peak in the house price index should indicate a peak in delinquency rates.

[1] The 90+ deliquency rate is the proportion of all receivables 90 or more days past in a given quarter

[2] The Mortgage Bankers Association is the largest US society representing the US real estate market, with over 2,400 members(banks, mortgage brokers, mortgage companies, life insurance companies, etc.)

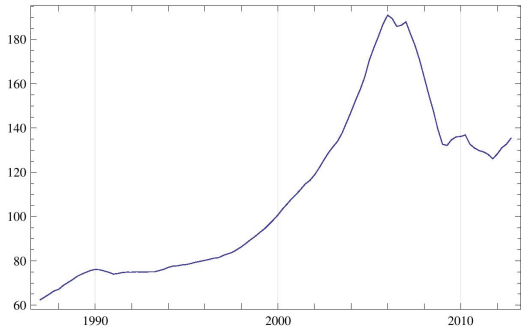


Figure 1 The house price index - underlying factor I_t

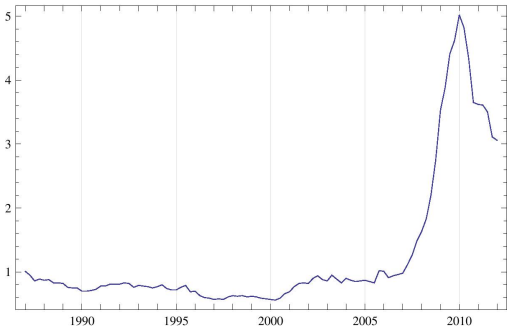


Figure 2 The US 90+ Delinquency Rates - factor DR_t

3.2 Estimation

We took default rate DR_t as the delinquency rate from the dataset; the factor I_t was taken as the house price index from the dataset. Then we evaluated ΔY_t according to Equation (5), where $\Psi(x) = \Phi(x)$ is a distribution function of a standard normal distribution. Then we evaluated the difference of I_t and $L_t = DR_t \cdot h(\Delta I_t)$. Finally, we were fitted our model.

In Table 1 and Table 2 we can see estimation of coefficients from Equation 12 and Equation 13 (all coefficients are significant) with standard deviation and p-value obtained by t-test.

	Coefficient	Standard dev.	p-value	
const	−0.0030	0.0026	0.2406	
ΔL_{t-3}	173.5630	32.5078	0.0000	***
ΔL_{t-4}	−46.4339	24.5574	0.0619	*
ΔI_{t-2}	0.0023	0.0009	0.0095	***
ΔY_{t-1}	−0.6236	0.0923	0.0000	***
ΔY_{t-2}	−0.6092	0.1011	0.0000	***

Table 1 Fitting of Equation 12 - dependent variable ΔY_t

4 Forecast

We have forecast the default rate DR_t, the loss given default $LGD_t = h(\Delta I_t)$ and the loss of the bank $L_t = h(\Delta I_t) \cdot DR_t$ for the following quarter, i.e., 2013Q3. The data for the default rate ends in 2012Q1 but the data for the house price index ends in 2012Q3. We constructed the forecast in two steps. In the first step we forecast the default rate up to 2012Q3 and in the second step we simultaneously forecast

	Coefficient	Standard dev.	p-value	
const	0.0989	0.1483	0.5066	
ΔDR_{t-3}	520.2640	183.2700	0.0056	***
ΔDR_{t-4}	−557.7900	192.9140	0.0048	***
ΔY_{t-2}	12.2806	5.5212	0.0286	**
ΔY_{t-3}	29.4992	8.7178	0.0011	***
ΔI_{t-1}	1.0413	0.0971	0.0000	***
ΔI_{t-2}	−0.3774	0.1370	0.0071	***
ΔI_{t-3}	0.2122	0.0989	0.0347	**

Table 2 Fitting of Equation 13 - dependent variable: ΔI_t

the house price index and the default rate. The forecast is shown in Figure 3. The default rate is ratio form Equation 1, values of LGD and Loss are under assumption that exposure at default is unit.

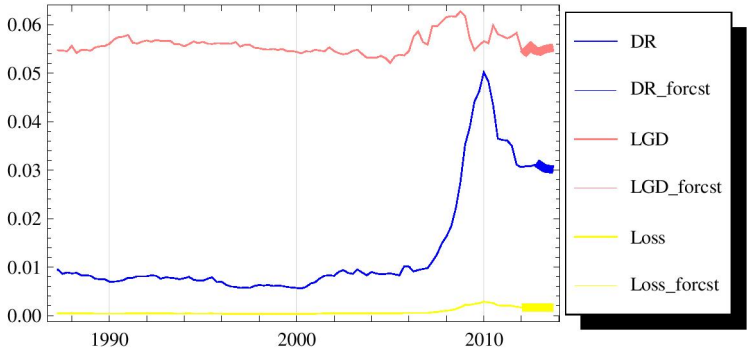


Figure 3 Forecast of DR_t, LGD_t and L_t for 2013Q3

5 Conclusions

We generalised [3] model, changed a linear sub-model into non-linear one, showed the statistical significance of non-linear dynamics. We applied our model to the real data and construct the forecast. We think that non-linearity is the key property of our model.

There are several topic for future research the main one is study of properties of functional AR process, especially existence of stationary distribution.

References

[1] Dullmann, K. and Trapp, M.: Systematic risk in recovery rates - an empirical analysis of U.S. corporate credit exposure. Working paper, Deutsche Bundesbank, Frankfurt, Germany.

[2] Frye, J. : Depressing recoveries. *Risk*,13(11): 106-111, 2000.

[3] Gapko, P. and Šmíd, M.: Dynamic Multi-Factor Credit Risk Model with Fat-Tailed Factors. Czech Journal of Economics and Finance, 62(2): 125–140, 2012.

[4] Park, Y. and Bang,D.: Bang. Loss given default of residential mortgages in a low ltv regime: Role of foreclosure auction process and housing market cycles. *Journal of Banking & Financeournal of Banking & Finance*, 39: 192–210, 2014.

[5] Pykhtin, MV : Unexpected Recovery Risk. *Risk*, 16(8):74–78, 2003.

[6] Vasicek, O.: The distribution of loan portfolio value. *RISK*, 15(12): 160 − 162, 2002.

[7] Qi, M. and Yang, X.: Loss given default of high loan-to-value residential mortgages. *Journal of Banking & Financeournal of Banking & Finance*, 33(5): 788–799, 2009.

[8] Qi, M.: *Credit Securitizations and Derivatives: Challenges for the Global Markets* , pages 33–52, Mortgage Credit Risk, 2013.

[9] Zhang, Y., Chi, L., Liu, F. and Ji, L.: Local Housing Market Cycle and Loss Given Default: Evidence from Sub-Prime Residential Mortgages. *International Monetary Fund.* Working paper, 2010.

Heuristic Algorithms for Multiple Messenger Problem with Multiple Depots

Jan Fábry[1]

Abstract. Multiple messenger problem is the generalization of messenger problem where more messengers can be used in the solution. In the paper, multiple depots are considered instead of one being shared by messengers. Because of NP-hardness of this problem, heuristic methods are used for real applications. To solve a static variant of multiple messenger problem, the nearest neighbor algorithm and insertion method are used. The paper also describes the exchange algorithm for the improvement of obtained solutions. The application, developed in VBA for MS Excel can solve large-scale problems with multiple messengers located in separated depots. Results of the computational experiments on real and generated data are presented in the paper.

Keywords: travelling salesman problem, multiple messenger problem, nearest neighbor algorithm, insertion method, exchange algorithm.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Messenger problem (MP) is analogous to dial-a-ride problem (DARP) [3]. While the DARP mostly concerns with the transport of handicapped or disabled persons, the MP deals with the transport of packages. In both variants, request of customers consists of a pick-up location and a delivery location. A static version of the problem assumes the knowledge of all customers in advance. De Paepe et al. [10] investigate the computational complexity of this class of problems. Cordeau [2] offers branch-and-cut algorithm for solving this problem. For the problem with time windows, Psaraftis [12] defined the exact algorithm. Because of NP-hardness of the problem, approximation algorithms are necessary to be used for large-scale instances. The heuristic algorithms for transportation of handicapped persons are presented by Toth and Vigo [14]. Jorgensen et al. [8] use genetic algorithms for solving the DARP. Use of evolutionary algorithms to solve distribution problems is shown in [5] and [9]. Multiple messenger problem (MMP) is based on multiple travelling salesperson problem (MTSP) described in [1]. Metaheuristic approach to solving the problems with multiple vehicles is applied by Cordeau and Laporte [4], or Song et al. [13]. In [6] heuristic algorithms for the MMP with all messengers in one depot are developed as the modifications of nearest neighbor method, insertion algorithm and exchange method. In the paper, those methods are modified for problems with multiple messengers located in multiple depots. Nearest neighbor algorithm and insertion method are used for generating routes and the exchange algorithm is applied for improvement of the solution. In fact, the proposed method is based on effective one-way transfers of shipments from certain route to another one. Time limitation for each route has to be respected. The paper presents computational experiments with all proposed methods on the combination of real minimal distance matrix and generated shipments.

2 Heuristic algorithms for generating multiple routes

In messenger problem, a location of pick-up and a location of delivery are given for each package. Suppose n packages have to be transported by messengers, all are known in advance. Let us denote each pick-up node in the network by the even number and each delivery node by the odd number. In addition, if the package is picked up at node i it will be delivered at node $i + 1$. For service, K messengers in separated depots are available. Let us assume, because of the proposed algorithms, K is an odd number. In case

[1]University of Economics Prague, Department of Econometrics, W. Churchill sq. 4, 130 67 Prague 3, fabry@vse.cz

the real number of messengers is even, a dummy messenger is introduced. Because of possibility of adding new customers in a dynamic version of the problem, nodes $1, 2, \ldots, K$ specify depots of messengers, nodes $K+1, K+3, \ldots, K+2n-1$ correspond to pick-up locations and $K+2, K+4, \ldots, K+2n$ denote delivery locations. Minimal distances c_{ij} and associated travel times d_{ij} $(i, j = 1, 2, \ldots, K+2n-1; i \neq j)$ for all pairs of locations are given. For generating routes, modifications of two heuristic algorithms are proposed: the nearest neighbor algorithm and the insertion method.

2.1 Nearest neighbor algorithm

The proposed algorithm, as the modification of the heuristics developed for the MTSP, is a sequential construction process that starts by initializing the current routes at all depots. Several rules can be applied for routes generation. Because we assume all available messengers will be used for packages transport, nearest locations to depots are selected to initiate all routes. Let A_k $(k = 1, 2, \ldots, K)$ is a set of locations that had not been visited yet and could be visited by messenger k at a given step. At the beginning, only pick-up locations can be assigned to each messenger, and therefore, all these sets contain only even numbers $K+1, K+3, \ldots, K+2n-1$. Let us denote S_k a sequence of locations being subsequently visited by messenger k and i_k is the last node, the messenger k visited. For each route, z_k is a distance which messenger k travels on the route. The total distance travelled by all messengers, is denoted z.

Algorithm 1

Step 1:

for $k = 1, 2, \ldots, K : A_k = \{K+1, K+3, \ldots, K+2n-1\}$

Step 2:

for $k = 1, 2, \ldots, K$ repeat

$c_{ks} = \min\limits_{j \in A_k} c_{kj}$; $S_k = \{k, s\}$; $z_k = c_{ks}$; $T_k = d_{ks}$; $A_k = A_k + \{s+1\}$; $B_k = \{s+1\}$; $i_k = s$

for $l = 1, 2, \ldots, K : A_l = A_l - \{s\}$

Step 3:

for $k = 1, 2, \ldots, K$ repeat

for all $j \in A_k$ repeat

$A'_k = A_k$; $B'_k = B_k$; $T'_k = T_k$; $r = j$

if j is even then $A'_k = A'_k + \{j+1\}$ and $B'_k = B'_k + \{j+1\}$

repeat while $B'_k \neq \emptyset$

$d_{rv} = \min\limits_{i \in B'_k} d_{ri}$; $T'_k = T'_k + d_{rv}$; $B'_k = B'_k - \{v\}$; $r = v$

$T'_k = T'_k + d_{rk}$; if $T'_k > Tmax$ then $A'_k = A'_k - \{j\}$

$c_{i_m s} = \min\limits_{k=1,2,\ldots,K} \min\limits_{j \in A'_k} c_{i_k j}$; $S_m = S_m + \{s\}$; $z_m = z_m + c_{i_m s}$; $T_m = T_m + d_{i_m s}$; $i_m = s$

if s is even then

for $k = 1, 2, \ldots, K : A_k = A_k - \{s\}$

$A_m = A_m + \{s+1\}$; $B_m = B_m + \{s+1\}$

else

$A_m = A_m - \{s\}$

Step 4:

if $A_k = \emptyset$ for all $k = 1, 2, \ldots, K$ then go to Step 5 else go to Step 3

Step 5:

for $k = 1, 2, \ldots, K : S_k = S_k + \{k\}$; $z_k = z_k + c_{i_k k}$; $z = \sum\limits_{k=1}^{K} z_k$

End

In Step 1, all possible pick-up nodes are set for all messengers. Step 2, in which the nearest nodes to depots are found in the ascending order of messenger numbers, can be modified. First, the pair of a messenger and a pick-up node with minimal distance is found and the corresponding route is initialized.

Then, for remaining messengers this selection is repeated until routes for all messengers are initialized. In fact, this rule is used in Step 3 to expand routes. Time limitation for routes is considered in the algorithm. Let $Tmax$ be time limit for each route, and T_k denotes current time travelled by messenger k. Referring to the notation given above, $T_k = \sum_{j=1}^{|S_k|-1} d_{u_j,u_{j+1}}$, where u_j is the location (node) number on the j-th position in the route. Apparently, $u_1 = k$ and $u_{|S_k|} = i_k$. Because there can be some non-visited delivery locations associated with the pick-up locations visited before, it is useful to denote B_k as the set of such delivery nodes. The reason is that for final travel time of messenger k it is necessary to include not only current time but also time spent by future visiting all delivery nodes, obligatory for the messenger. In Step 3, node $j \in A_k$ can be available for the selection as the nearest neighbor to messenger k only in case it is possible to finish the route within time $Tmax$. If even node j is selected it is necessary to include also node $j + 1$ to the set of obligatory nodes. The algorithm does not consider the situations that all messengers are unable to accept the package into their route because of limited time. Therefore, using this heuristic approach, it is necessary to refuse any customers and to put over this requirement to next day. Of course, it is possible to change the algorithm or modify some respected rules for this purpose. Nevertheless, even the application of the optimization model does not assure that all requirements will be satisfied.

2.2 Insertion method

Insertion algorithm is based on inserting locations to current routes what extends their length. If the insertion method is used for TSP the route is obviously initialized by finding the farthest location s from the depot 1. Then, the cycle $\{1, s, 1\}$ is created as the initial route. Another location is inserted to the cycle with the objective to minimize its extension. This approach is repeated until all locations are incorporated in the Hamiltonian cycle. In case of multiple messenger problem, two important rules have to be respected: pick-up location and related delivery location have to be inserted to the same route, and pick-up location has to be inserted to the route in front of related delivery location. In addition, delivery location can be inserted exactly behind related pick-up one or behind several locations further [7]. If multiple messengers are considered in the problem, finding the farthest location to each depot in the initialization step can be disadvantageous due to the farthest location to the depot can be very close to another one. In such situation it would be inefficient to include the location in the route of the concerned messenger. Therefore, the different approach should be be used.

In the following algorithm we are finding, for each messenger k, the location that is closer to depot k than to all other depots. If more locations satisfy this condition, the farthest one to depot k determines the initial route (see variables $LowDist$ and $LowLoc$ in the proposed algorithm). If no such location exists, i.e. all locations are farther to depot k than to any of remaining depots, as the initial location, it is taken the one with the minimal difference between its distance from depot k and minimal distance from other depots (see variables $UpDif$ and $UpLoc$ below). Referring to previous notation, A_k $(k = 1, 2, \ldots, K)$ is a set containing locations that can be visited by messenger k at a given step. At the beginning, both pick-up and delivery locations can be selected to initiate cycles. If S_k is a sequence of locations being subsequently visited by messenger k, $u_i^k \in S_k$ is the node number on position i of sequence S_k. Finally, let us denote $h_k = |S_k|$.

In each iteration, the related pick-up and delivery locations are inserted. Two possibilities for insertions are acceptable [7]: delivery location will be visited immediately after visiting the pick-up location, or delivery location will be visited later, after visiting other locations. In the first case, the value of $\Delta zImm_{rj}$ expresses the minimal extension of the current route of messenger k, if pick-up node r and delivery node $r+1$ are inserted together behind the node on position j in S_k. Value of $\Delta tImm$ is calculated to check whether it is acceptable to include those locations in the route due to time limitation. Similarly in the second case, the value of $\Delta zLat_{rvw}$ corresponds to the minimal extension of the route, if node r is inserted behind the node on position v and $r + 1$ later, behind the node on position w. Variable $\Delta tLat$ is used to check feasibility in terms of time limitation. Finding all acceptable insertions minimizing the extensions of individual routes, it is realized the most advantageous insertion of nodes $Pick$ and $Pick+1$ in the route of messenger Msg behind position $After1$ in S_{Msg} (if $Imm = true$), or behind positions $After1$ and $After2$ (if $Imm = false$). All sets A_k are modified finally. Similarly to Algorithm 1, the proposed insertion algorithm does not consider the situation when any package cannot be accepted by all messengers due to time limitations. The algorithm has to be modified for those instances.

Algorithm 2

Step 1:

for $k = 1, 2, \ldots, K : A_k = \{K+1, K+2, \ldots, K+2n\}$

Step 2:

for $k = 1, 2, \ldots, K$ repeat

$LowDist = -\infty$; $UpDif = +\infty$

for all $j \in A_k$ repeat

if $c_{kj} \leq \min\limits_{\substack{l=1,2,\ldots,K \\ l \neq k}} c_{lj}$ then

if $c_{kj} > LowDist$ then

$LowDist = c_{kj}$; $LowLoc = j$

else if $(c_{kj} - \min\limits_{\substack{l=1,2,\ldots,K \\ l \neq k}} c_{lj}) < UpDif$ then

$UpDif = (c_{kj} - \min\limits_{\substack{l=1,2,\ldots,K \\ l \neq k}} c_{lj})$; $UpLoc = j$

if $LowDist > -\infty$ then $s = LowLoc$ else $s = UpLoc$

if s is even then

$S_k = \{k, s, s+1, k\}$; $z_k = c_{ks} + c_{s,s+1} + c_{s+1,k}$; $T_k = d_{ks} + d_{s,s+1} + d_{s+1,k}$

else

$S_k = \{k, s-1, s, k\}$; $z_k = c_{k,s-1} + c_{s-1,s} + c_{sk}$; $T_k = d_{k,s-1} + d_{s-1,s} + d_{sk}$

$h_k = 4$;

for $l = 1, 2, \ldots, K$ repeat

$A_l = A_l - \{s\}$; if s is even then $A_l = A_l - \{s+1\}$ else $A_l = A_l - \{s-1\}$

Step 3:

$\Delta z = \infty$

for $k = 1, 2, \ldots, K$ repeat

for all even $r \in A_k$ repeat

$\Delta zImm_{rj} = \min\limits_{i=1,2,\ldots,h_k-1} (c_{u_i^k r} + d_{r,r+1} + c_{r+1,u_{i+1}^k} - c_{u_i^k u_{i+1}^k})$

$\Delta tImm = d_{u_j^k r} + d_{r,r+1} + d_{r+1,u_{j+1}^k} - d_{u_j^k u_{j+1}^k}$

$\Delta zLat_{rvw} = \min\limits_{\substack{i=1,2,\ldots,h_k-2 \\ m=i+1,i+2,\ldots,h_k-1}} (c_{u_i^k r} + c_{r u_{i+1}^k} - c_{u_i^k u_{i+1}^k} + c_{u_m^k,r+1} + c_{r+1,u_{m+1}^k} - c_{u_m^k u_{m+1}^k})$

$\Delta tLat = d_{u_i^k r} + d_{r u_{i+1}^k} - d_{u_i^k u_{i+1}^k} + d_{u_m^k,r+1} + d_{r+1,u_{m+1}^k} - d_{u_m^k u_{m+1}^k}$

if $T_k + \Delta tImm < Tmax$ and $\Delta zImm_{rj} < \Delta z$ then

$Imm = true$; $Msg = k$; $Pick = r$; $After1 = j$; $\Delta z = \Delta zImm_{rj}$; $\Delta T = \Delta tImm$

if $T_k + \Delta tLat < Tmax$ and $\Delta zLat_{rvw} < \Delta z$ then

$Imm = false$; $Msg = k$; $Pick = r$; $After1 = v$; $After2 = w$; $\Delta z = \Delta zLat_{rvw}$; $\Delta T = \Delta tLat$

if Imm and $\Delta z < \infty$ then

insert nodes $Pick$ and $Pick + 1$ to the sequence S_{Msg} to positions $After1 + 1$ and $After1 + 2$

else if $\Delta z < \infty$ then

insert nodes $Pick$ and $Pick + 1$ to the sequence S_{Msg} to positions $After1 + 1$ and $After2 + 1$

if $\Delta z < \infty$ then

$h_{Msg} = h_{Msg} + 2$; $z_{Msg} = z_{Msg} + \Delta z$; $T_{Msg} = T_{Msg} + \Delta T$

for $k = 1, 2, \ldots, K$ repeat

$A_k = A_k - \{Pick\}$; $A_k = A_k - \{Pick + 1\}$

Step 4:

if $A_k = \emptyset$ for all $k = 1, 2, \ldots, K$ then go to Step 5 else go to Step 3

Step 5:

$$z = \sum_{k=1}^{K} z_k$$

End

3 Exchange algorithm for improving generated routes

Both nearest neighbor algorithm and insertion method described above generate routes starting at different depots. This solution can be improved using another heuristic approach. For the MMP with messengers starting in one depot such heuristics was proposed in [6]. In this paper, the method is modified for routes starting in multiple depots. The main idea is to transfer some shipments from their original routes to other ones if it is advantageous. The transfer of the shipment consists in its excluding from generated route to another one. Only exchanges respecting given time limit can be accepted. Thus, the algorithm is based on the generated routes using Algorithm 1 or Algorithm 2.

Let K be the number of generated routes given by sequences S_k $(k = 1, 2, \ldots, K)$. Let $h_k = |S_k|$ and $u_i^k \in S_k$ is the location on position i of the sequence $(i = 1, 2, \ldots, h_k)$. Values z_k and T_k represent the length and travel time of route k. There are many modifications of proposed algorithm. The transfer approach is offered in the simplified form.

In the first step, it is found, for each route, a shipment (i.e. even node i and odd node $i + 1$), which brings maximal reduction $\Delta zExcl_k$ $(k = 1, 2, \ldots, K)$ of the length of the route when it is excluded. These shipments are candidates for possible transfer to another route. Then, route m with the maximal reduction $\Delta zExcl_m$ is selected. Let us assume that this reduction is achieved for nodes r and $r + 1$ on positions v and w in S_m. In the second step, the insertion algorithm described above is applied for the shipment determined in the previous step. If messenger Msg is found with the minimal extension $\Delta zExt_{Msg}$ of the route after the acceptable (in terms of limited time) insertion of nodes r and $r+1$, such that $\Delta zExt_{Msg} < \Delta zExcl_m$, the transfer of shipment from messenger m to messenger Msg is realized. Sequences S_m and S_{Msg} are changed and all related values recalculated: h_m, h_{Msg}, z_m, z_{Msg}, T_m and T_{Msg}. These two steps are repeated while effective transfers of shipments are possible.

Several corrections and modifications of proposed method can be made. In case there is no possible insertion because of limited time, nodes r and $r + 1$ are moved to "tabu" list and the first step is re-executed with the elimination of those shipments from possible transfers. If no effective transfer is possible, we can release the rule of finding the maximal reduction and minimal extension, because the total impact of the change is important. Another possibility is to implement the effective exchange of shipments instead of one-way transfer.

4 Computational experiments

In [11] several experiments were executed with the application developed in VBA for MS Excel. Testing instances were divided into four groups according to a number of messengers and a number of packages: cases with 3, 5, 7 and 11 messengers with 10, 20, 30 and 40 packages. In each group, 10 instances were generated using the matrix of real minimal distances $c_{ij} \in [5, 60]$ $(i, j = 1, 2, \ldots, K + 2n - 1; i \neq j)$ between 650 locations. For the purpose of the simplification of experiments, the assumption of $d_{ij} = c_{ij}$ was introduced and time limits were set to 120 units in case of 3 and 5 messengers and 150 units for instances with 7 and 11 messengers.

Insertion method (Algorithm 2) provides better solution than nearest neighbor algorithm (Algorithm 1) in 75% of generated instances. Exchange algorithm was applied for potential improvement to all solutions. In case of nearest neighbor algorithm, 70% of obtained solutions were improved, while in case of insertion method only 15% of solutions were improved. Improving the results, 65% of all solutions were finally better for insertion method than nearest neighbor algorithm. The improvement leads to decrease of total length of all routes by approximately 5% in case of the nearest neighbor algorithm in instances with 3 and 5 messengers, and by 1% in instances with 7 and 11 messengers. Improvement of the solution in case of insertion method leads to decrease by approximately 1% in all groups.

5 Conclusions and future work

In the paper, modifications of two heuristic algorithm are proposed to generate routes in multiple messenger problem with separated depots: nearest neighbor algorithm and insertion method. In both approaches, a static version of messenger problem with time limitation is considered. For improvement of generated solutions, the exchange algorithm is proposed. As the results of computational experiments show, the improvement is significant only in case of nearest neighbor algorithm, while insertion method

provides effective solution itself. Unfortunately, because of NP-hardness of the problem, results of experiments cannot be compared with the optimal solutions of even small-scale instances. In additions, real travel times should be found to make the computational experiment more predicative.

Future research will be aimed at the modifications of all proposed algorithms mentioned above, especially in case of the exchange algorithm, where more approaches can be realized. In case of time limitations, messenger problem can be extended to instances with time windows. If large packages are delivered, it is necessary to consider capacities of all messengers. Finally, all modifications can be applied for dynamic versions of intended problems.

Acknowledgements

This research was supported by the project F4/11/2013 founded by the Internal Grant Agency of the University of Economics Prague.

References

[1] Bektas, T.: The Multiple Traveling Salesman Problem: an Overview of Formulations and Solution Procedures, *Omega* **34** (2006), 209–219.

[2] Cordeau, J.-F.: A Branch-and-Cut Algorithm for the Dial-a-Ride Problem, *Operations Research* **54** (2006), 573–586.

[3] Cordeau, J.-F. and Laporte, G. A.: The Dial-a-Ride Problem: Models and Algorithms, *Annals of Operations Research* **153** (2007), 29–46.

[4] Cordeau, J.-F. and Laporte, G. A.: A Tabu Search Heuristics for the Static Multi-Vehicle Dial-a-Ride Problem, *Transportation Research Part B* **37** (2003), 579–594.

[5] Čičková, Z., Brezina, I. and Pekár, J.: A Memetic Algorithm for Solving the Vehicle Routing Problem. In: *Proceedings of Mathematical Methods in Economics '11*, Janska Dolina, 2011, 125–128.

[6] Fábry, J. and Kobzareva, M.: Multiple Messenger Problem. In: *Proceedings of Mathematical Methods in Economics '12*, Karviná, 2012, 141–147.

[7] Fábry, J.: Dynamic Messenger Problem, *Communications* **4** (2007), 66–69.

[8] Jorgensen, R. M., Larsen, J. and Bergvinsdottir, K. B.: Solving the Dial-a-Ride Problem using Genetic Algorithms, *Journal of the Operational Research Society* **58** (2006), 1321–1331.

[9] Melechovský, J.: Evolutionary Local Search Algorithm to Solve the Multi-Compartment Vehicle Routing Problem with Time Windows. In: *Proceedings of Mathematical Methods in Economics '12*, Karviná, 2012, 564–568.

[10] de Paepe, W. E., Lenstra, J. K., Sgall, J., Sitters, J. A. and Stougie, L.: Computer-Aided Complexity Classification of Dial-a-Ride Problems, *INFORMS Journal on Computing* **16** (2004), 120–132.

[11] Přibylová, L.: *Heuristiky pro kapacitní úlohy kurýrní služby.* Diploma thesis, University of Economics Prague, 2014.

[12] Psaraftis, H. N.: An Exact Algorithm for the Single Vehicle Many-to-Many Dial-a-Ride Problem with Time Windows, *Transportation Science* **17** (1983), 351–357.

[13] Song, J., DePuy, G. W and Evans, G. ,W.: A Meta-euristic for the Multi-Vehicle Dial-A-Ride Problem with Time Windows. In: *Proceedings of IIE Annual Conference '04*, Houston, 2004, 1–6.

[14] Toth, P. and Vigo, D.: Heuristic Algorithms for the Handicapped Persons Transportation Problem, *Transportation Science* **31** (1997), 60–71.

Compositional tables analysis with application to manufacturing

Kamila Fačevicová[1], Valentin Todorov [2], Karel Hron[3]

Abstract. Compositional tables as a special case of compositional data and continuous counterpart of contingency tables carry relative information about relationship between row and column factors. Consequently, cells of a compositional table contain quantitatively expressed relative contributions on a whole, thus for their statistical analysis only ratios between cells are informative. The standard Euclidean geometry is not appropriate for this purpose and should be replaced by the Aitchison geometry on the simplex that fully accounts for specific features of compositional tables. Moreover, the Aitchison geometry enables to decompose the compositional table into its independent and interactive parts; in particular, the latter one is crucial for further investigation of the relationship between row and column factors. Furthermore, by assigning orthonormal coordinates to decomposition of compositional tables, standard methods for their statistical processing can be applied. The aim of the contribution is to study the relationship between resource efficiency and technology in manufacturing. For this purpose, a sample of 2×3 compositional tables was collected and the interaction tables were further analyzed using the compositional tables methodology. It was shown that any choice of interpretable orthonormal coordinates of the interaction tables leads to the same general conclusions concerning the analyzed relationship.

Keywords: compositional data, isometric logratio transformation, manufacturing.

JEL classification: C44
AMS classification: 90C15

1 Introduction

The relationship between two factors is usually analyzed using contingency tables, which cells result from a random sample of variables with multinomial distribution. This basic problem can be further extended, when a sample of tables, whose cells have in general continuous nature, is considered. In this paper we will study the relation between resource efficiency and technology level in manufacturing based on a sample of tables carrying information on these two factors. Increasing resource efficiency is a key factor which drives down costs and improves competitiveness thus securing growth and jobs. To understand the performance of industry and to improve its resource efficiency it is necessary to study the relative importance of one kind of industrial activity with respect to the others and how this structure changes (see [8]). Value added (the difference between output and intermediate consumption for a given industry) is a primary indicator of economic performance and the ratio of value added to output can be used as a measure of resource efficiency. The second factor is the classification of the manufacturing sectors by technological intensity. The sample contains data for 77 countries in 2008, collected through annual industrial survey or periodic census by the national statistical offices and then provided to UNIDO (United Nations Industrial Development Organization) in order to update its database and disseminate the data internationally at the 4-digit level of the International Standard Industrial Classification of All

[1]Palacký University, Department of Mathematical Analysis and Applications of Mathematics, Department of Geoinformatics, 17. listopadu 12, CZ-77146 Olomouc, Czech Republic, kamila.facevicova@gmail.com
[2]United Nations Industrial Development Organization (UNIDO), Vienna International Centre, P.O. Box 300, 1400 Vienna, Austria, valentin.todorov@chello.at
[3]Palacký University, Department of Mathematical Analysis and Applications of Mathematics, Department of Geoinformatics, 17. listopadu 12, CZ-77146 Olomouc, Czech Republic, hronk@seznam.cz

CZE	HT	MT	LT
VA	$1.87814 \cdot 10^{10}$	$1.50362 \cdot 10^{10}$	$1.03281 \cdot 10^{10}$
I	$7.46011 \cdot 10^{10}$	$4.67950 \cdot 10^{10}$	$2.95888 \cdot 10^{10}$

CZE	HT	MT	LT
VA	0.09625	0.07706	0.05293
I	0.38231	0.23981	0.15164

Table 1 Relative structure of manufacturing production in Czech Republic in 2008 according to the value added-input structure (resource efficiency) and level of technology intensity in absolute values (left table) and in proportions (right table).

Economic Activities (ISIC) Revision 3 [8]. Each table in this sample could be considered a compositional table (a special case of D-part compositional data [1, 7]), since the whole information about relation between resource efficiency and technology intensity is contained in ratios between the parts. This is also the typical property of compositional tables and compositional data in general.

2 Compositional tables and their geometry

$I \times J$ compositional table $\mathbf{x}$ represents a special case of D-part compositional data, which carries relative information about relationship between row and column factors with I and J values, respectively. Its parts are strictly positive and describe quantitatively their relative contributions on the whole [2]. Thus the absolute values of the parts are not important, since all the relevant information in the composition is contained in the ratios between parts. Using the closure operation $\mathcal{C}(\mathbf{x})$, each composition can be rescaled to a prescribed constant sum constraint κ, like $\kappa = 1$ in case of proportions.

The analyzed sample consists of 2×3 compositional tables, since factor *resource efficiency* has values Value added (VA) and Input (I) and factor *technology level* has three values, Medium-high and High technology, Medium-low technology and Low technology (HT, MT and LT). Example of one such table, both in original values and in proportions, is provided in Table 1.

The sample space of representations of $I \times J$-part compositional tables is the $(I \cdot J)$-part simplex, and their geometrical features are captured by the Aitchison geometry [1, 4]. This geometry is represented by operations of perturbation and power transformation. Components of perturbation of two $I \times J$ compositional tables, $\mathbf{z} = \mathbf{x} \oplus \mathbf{y}$, are $z_{ij} = (x_{ij}y_{ij})/\sum_{rs} x_{rs}y_{rs}$ (already expressed in their proportional representation). Powering by a real constant α results in $I \times J$ compositional table $\mathbf{z} = \alpha \odot \mathbf{x}$ with parts $z_{ij} = x_{ij}^{\alpha}/\sum_{rs} x_{rs}^{\alpha}$.

3 Compositional tables analysis

As in the case of contingency tables, the goal of compositional tables analysis is primarily to study the relationship between row and column factors and to answer the question, whether these factors are independent, or not. The analysis of independence is simplified substantially through the decomposition of the original table onto its independent and interactive parts [3, 6]

$$\mathbf{x} = \mathbf{x}_{ind} \oplus \mathbf{x}_{int}. \tag{1}$$

The *independence table* $\mathbf{x}_{ind}$ carries relative information about ratios between rows or columns, respectively, and has parts $x_{ij}^{ind} = \left(\prod_{k=1}^{I}\prod_{l=1}^{J} x_{kj}x_{il}\right)^{\frac{1}{IJ}}$. The information about ratios between rows and columns (interactions in the table) is carried by the second part of decomposition (1), the *interaction table* with entries $x_{ij}^{int} = \left(\prod_{k=1}^{I}\prod_{l=1}^{J} \frac{x_{ij}}{x_{kj}x_{il}}\right)^{\frac{1}{IJ}}$. In the case of independence between factors in a compositional table, the whole information about the original table $\mathbf{x}$ is contained in the independence table $\mathbf{x}_{ind}$ and the interaction table $\mathbf{x}_{int}$ equals just to the neutral element $\mathbf{n}$ (all its entries are equal). On the other hand, the importance of the interaction table will increase, if the relationship between factors becomes stronger.

The interaction tables for all 77 countries in the sample have been calculated. The mean interaction table (2), resulting from analysed data set, indicates that there is independence between resource efficiency and technology intensity, since there are not big differences between its parts and thus it could approach

Coordinate	Analytical form	Sample mean	Sample st. dev.
z_{22}^{int}	$\frac{1}{2}\ln\frac{x_{11}x_{22}}{x_{12}x_{21}}$	0.1165	0.3060
z_{23}^{int}	$\frac{1}{\sqrt{12}}\ln\frac{x_{11}x_{12}x_{23}^2}{x_{21}x_{22}x_{13}^2}$	−0.0203	0.2052

Table 2 Sample mean and standard deviation of nonzero coordinates of the interaction table.

the neutral element. This preliminary conclusion should be further verified through the analysis in orthonormal (ilr) coordinates [4].

$$\overline{\mathbf{x}}_{int} = \begin{pmatrix} 0.16843 & 0.15613 & 0.17543 \\ 0.16454 & 0.17749 & 0.15797 \end{pmatrix} \tag{2}$$

Following [6], the interaction table can be expressed in ilr coordinates using formula

$$z_{rs}^{int} = \frac{1}{\sqrt{r \cdot s \cdot (r-1) \cdot (s-1)}} \ln \prod_{i=1}^{r-1} \prod_{j=1}^{s-1} \frac{x_{ij}x_{rs}}{x_{is}x_{rj}} \tag{3}$$

for $r = 2, 3, \ldots, I$ and $s = 2, 3, \ldots, J$. Generally, this formula leads to $(I-1)(J-1)$ nonzero orthonormal coordinates, here we get two such coordinates. When compositional data are expressed in the ilr coordinates, all standard statistical methods can be applied (see [4] for details). Thus, when we assume, that row and column factors are independent (i.e., the interaction table is neutral element), we can equivalently assume that its vector of ilr coordinates $\mathbf{z}_{int}$ equals to null vector. In case of random sample, this preliminary conclusion could be further verified through an appropriate test on zero values of both coordinates.

Note that for the case of analyzing the relationship between resource efficiency and technology in manufacturing, introduced originally in [6] with reverse order of HT, MT, LT columns, three systems of such orthonormal coordinates (up to orientation of axes) could be obtained by simply permuting cells in (3). One of aims of the paper is thus to demonstrate that even for any other choice of coordinates, like those presented here, the general conclusions about the analyzed relationship always remain the same.

4 Results and interpretation

Unfortunately, the sample which is formed by tables carrying information about distribution of the manufacturing production cannot be considered as a random sample due to possible spatial dependence. Thus, the hypothesis that the random vector $\mathbf{z}_{ind}$ equals to null vector is analyzed only descriptively through mean values and standard deviations of coordinates. Firstly, following the methodology presented above, all interaction tables have been expressed in coordinates. Nevertheless, for the analysis of the relationship between factors, only values of two nonzero coordinates of the interaction tables are sufficient; their sample means and standard deviations are collected in Table 2. Since both sample means are near zero and standard deviations are quite large comparing to the means, the corresponding theoretical interaction table would approach the neutral element and thus the resource efficiency and technology can be considered to be rather independent.

Both coordinates could be interpreted in terms of odds ratios (see [5, 6] for details). The first coordinate is connected with the odds ratio in the 2×2 table with higher and medium technology. The upper odd thus compares value added with input in high technology and the odd in denominator compares value added with input again, but for medium technology. Since the sample mean of z_{22}^{int} equals 0.1165, this odds ratio results in $e^{2\cdot(0.1165)} \doteq 1.26$ in average; the odd in nominator is thus slightly higher compared to the odd in denominator of the odds ratio.

The second coordinate is formed by the sum of two logarithms of odds ratios in the whole 2×3 table. The first odds ratio compares odd between value added and input in high technology (nominator) and odd between these two components for low technology (denominator). The second odds ratio compares similar odds for medium and low technology. The sum of logarithms of these two ratios is $2\sqrt{3} \cdot (-0.0203) = -0.0703 < 0$, thus, at least in one of the odds ratios the odd of value added at the low technology

is slightly higher than the corresponding odd for the higher technology. This coordinate could be also interpreted as $1/\sqrt{3}$ of logarithm of odds ratio which compares value added to input ratio for higher and medium technologies, aggregated by geometric mean, and the same ratio for low technology. This odds ratio equals 0.9655, thus we can conclude again about independence between factors for these two "new" levels of technology.

Possible deviations from the independence case in the set of countries contained in the sample can be also analyzed directly using values of coordinates of their interaction tables. In case of 2×3 compositional tables this results simply in a planar graph, displayed in Figure 1.

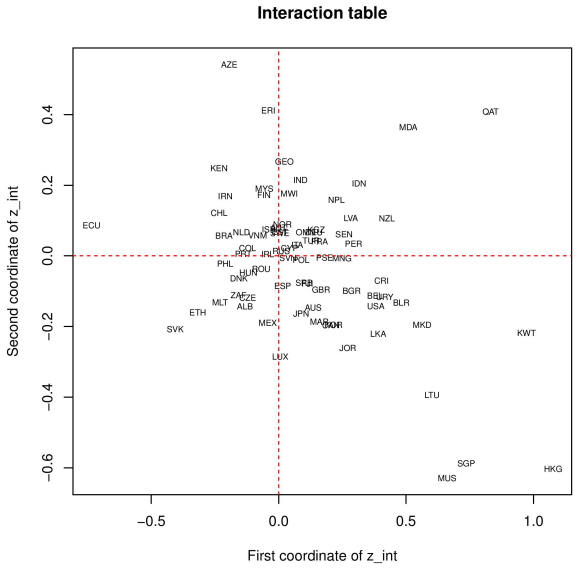


Figure 1 Scatterplot of coordinates of the interaction table for 77 countries.

It is easy to see that almost all countries are concentrated near to origin, what confirms our preliminary conclusion about independence between factors. However, there are also some outliers with quite high absolute value of the first or second coordinate (as Ecuador or Azerbaijan) or with both coordinates quite far from zero (Hong Kong). Since Ecuador has high negative value of the first coordinate and the second coordinate is approximately zero, the ratio of value added to input is much higher for medium technology than for high technology. In Figure 2 it is shown that the sector dominating the economics of Ecuador, with almost 30%, is *Coke, refined petroleum products, nuclear fuel*, part of medium technology and in Figure 2 (right panel) we see that the value added to output ratio of this sector is the highest in Ecuador. On the other hand, for *Chemicals and chemical products*, the most important sector of higher technology in this country, the value added to output ratio is quite low, which explains why Ecuador appears on left margin of the graph. On the other hand, in Azerbaijan the resource efficiency and technology intensity are independent when only higher and medium technology are of the interest. The independence is broken when the lower technology is taken into account. Figure 3 shows that almost all sectors of lower technology have low value added to input ratio, like *Food and beverages*, the second most important sector in Azerbaijan, which has almost the second lowest value of the value added to input ratio. Since this ratio is higher for sectors from medium and higher technology, the value of the second coordinate is high and Azerbaijan appears on upper margin of the graph.

Another interesting outlier is Hong Kong, which is characterized by quite high absolute values of both coordinates. As seen from Figure 4, the dominating sectors are *Food and beverages* and *Paper and paper products*, parts of low technology level and part of high technology *Machinery and equipment n.e.c.*, which are also the most resource efficient sectors in Hong Kong. Since sectors of medium technology level are not very important in this country, the value added to input ratio for high technology dominates the ratio for the medium technology and the first coordinate has quite high value. Consequently, the importance of the lower technology together with its high resource efficiency leads to high negative value of the second coordinate and places Hong Kong at the lower right corner of the graph.

Independence within the whole table (for all three technology levels) holds only for countries almost

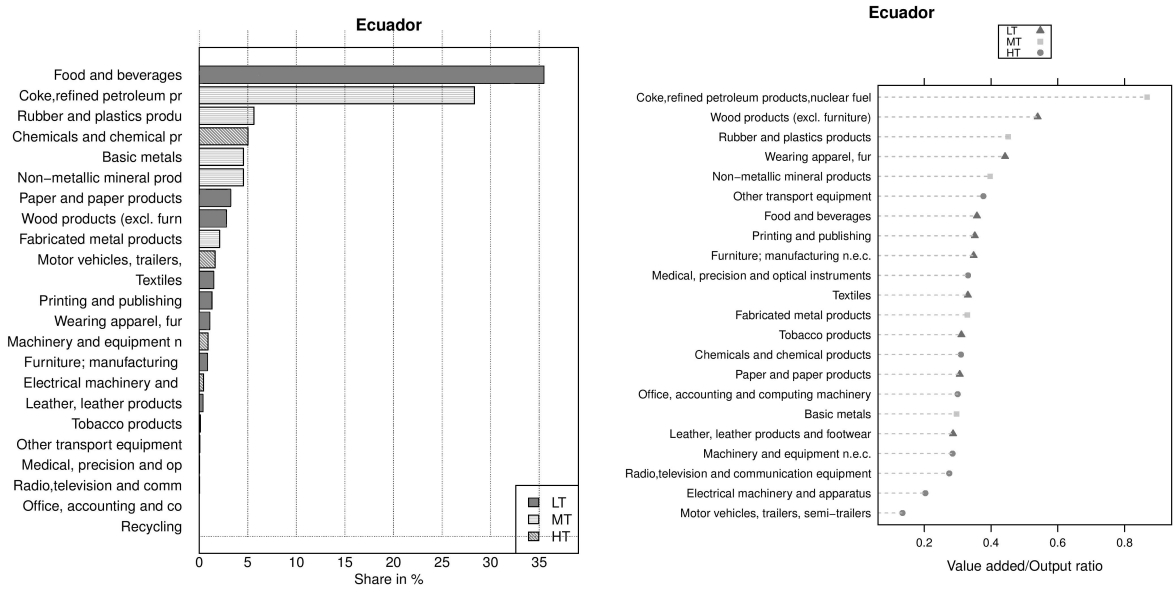


Figure 2 Share of sector value added in total value added and value added to output ratio for Ecuador.

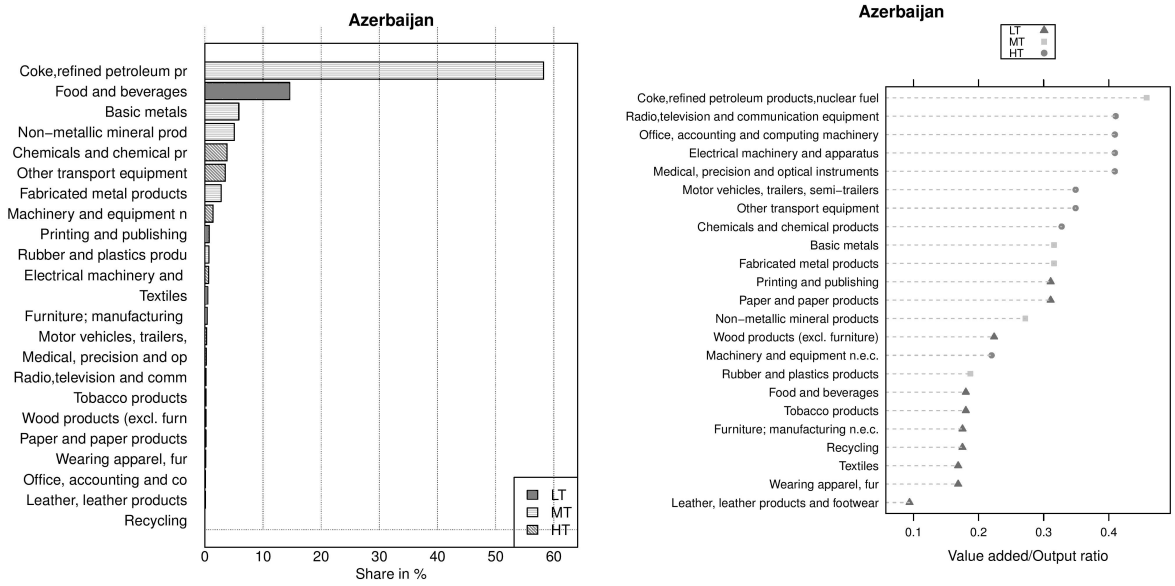


Figure 3 Share of sector value added in total value added and value added to output ratio for Azerbaijan.

at the origin (like Russia and Slovenia).

5 Conclusions

This paper deals with the problem of independence between resource efficiency and technology intensity in manufacturing. Because of the continuous nature of the data, the compositional approach has been used for the analysis of sample of tables. The method consists in decomposition of each table into independent and interactive parts and expression of interaction tables in coordinates. Since the mean values of the coordinates tend to be zero, we can conclude that there is independence between resource efficiency and technology intensity, nevertheless, detailed inspection of single coordinates provides even further valuable information. The coordinate representation also allows a graphical representation of data, which helps to identify countries that most deflect from the case of independence like Ecuador, Kuwait and Qatar. Finally, following the previous results, we can state that the above general conclusions could be obtained by using coordinate representation (3) with an arbitrary permutation of parts of the original tables.

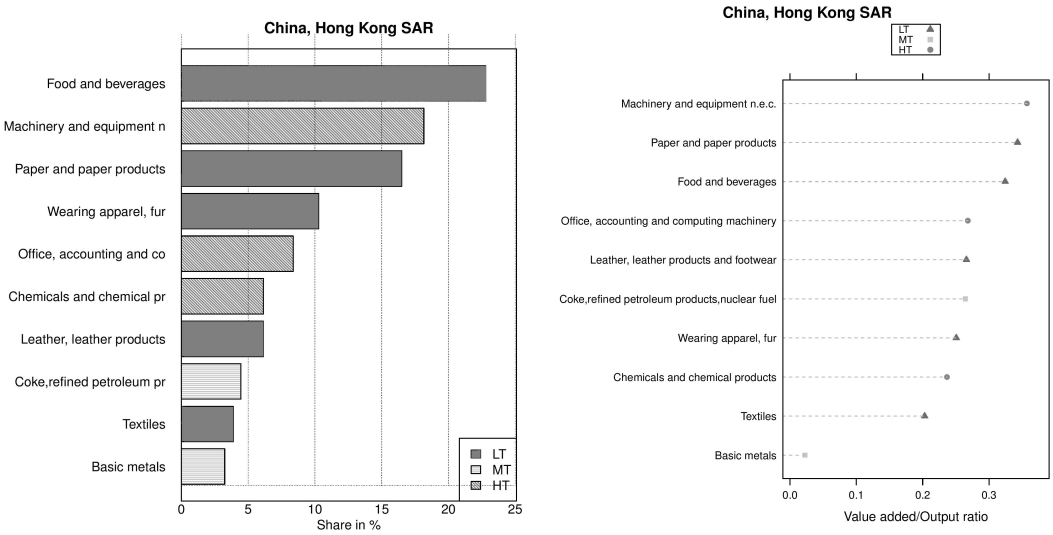


Figure 4 Share of sector value added in total value added and value added to output ratio for Hong Kong.

Disclaimer The views expressed herein are those of the authors and do not necessarily reflect the views of the United Nations Industrial Development Organization.

Acknowledgements

Supported by the Operational Program Education for Competitiveness - European Social Fund (project CZ.1.07/2.3.00/20.0170 of the Ministry of Education, Youth and Sports of the Czech Republic) and the grant IGA PrF_2014_028 Mathematical Models of the Internal Grant Agency of the Palacký University in Olomouc.

References

[1] Aitchison, J.: *The statistical analysis of compositional data.* Chapman and Hall, London, 1986.

[2] Egozcue, J. J.: Reply to "On the Harker Variation Diagrams" by J.A. Cortés. *Math Geosci* **41** (2009), 829–834.

[3] Egozcue, J. J., Díaz-Barrero, J. L. and Pawlowsky-Glahn, V.: Compositional analysis of bivariate discrete probabilities. In *Proceedings of CODAWORK'08, The 3rd Compositional Data Analysis Workshop.* (eds J. Daunis-i-Estadella and J. A. Martín-Fernández)University of Girona, Spain, 2008.

[4] Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barceló-Vidal, C.: Isometric logratio transformations for compositional data analysis. *Math Geol* **35** (2003), 279–300.

[5] Fačevicová, K., Hron, K., Todorov, V., Guo, D. and Templ, M.: Logratio approach to statistical analysis of 2×2 compositional tables. *Journal of Applied Statistics* **41** (2014), 944–958.

[6] Fačevicová, K., Hron, K., Todorov, V. and Templ, M.: Compositional tables analysis in coordinates. Submitted (2014).

[7] Pawlowsky-Glahn, V. and Buccianti, A.: *Compositional data analysis: Theory and applications.* Wiley, Chichester, 2011.

[8] Rohna, H., Pastewskib, N., Lettenmeierc, M., Wiesenc, K. and Biengec K.: Resource efficiency potential of selected technologies, products and strategies. *Science of The Total Environment* **473-474** (2014), 32-35.

[9] UN: International Standard Industrial Classification of All Economic Activities (ISIC) Rev. 3.1. (2002). (Available from http://unstats.un.org/unsd/cr/registry/regdnld.asp?Lg=1.)

The real use of the operations research methods in conjunction with logistics technologies

Anastasiya Fesenko

Abstract. This article focuses on the real use of operations research methods in conjunction with logistics technologies. The aim is to show the way of optimization and simulation methods application and to evaluate if these methods are suitable tools for logistics technology application. Large scale of those systems is used in the logistics and delivery services. Their main objectives are: to increase the efficiency of whole enterprise, avoid time and profit losses, to improve production capacity and as much as possible comply with customer requests. Technologies can be divided in: classic, telematic, virtual and complex. The group of complex logistic systems is comprised from Just in Time, Kanban, Quick Response, and Efficient Consumer Response. This paper is concerned mainly with this group of procedures. In the work were used mathematical tools including: methods of multi-criteria evaluation, mixed linear programming and simulation models. On the basis of mathematical models two examples of new technologies introduction and analysis of already functioning systems have been solved. The achieved results were compared to each other and their validity was analyzed.

Keywords: logistic technologies, JIT, Kanban, computer simulation, multi-criteria evaluation of alternatives, mixed linear programming.

JEL Classification: C44
AMS Classification:90C15

1 Introduction

In face of the challenges of global competition, business firms are concentrating more on the needs of customers and seeking ways to reduce costs, improve quality and meet the ever-rising expectation of their customers. To these ends, many of them have identified logistics as an area to build cost and service advantages [6]. Under application of logistics fall not only transportation of goods, but also the production, shipping, receiving, storage, consolidation, packaging, even sales of goods and management of all these areas. On the basis of [5] logistics involves the delivery of products or services for the client with assured quality and quantity. The logistics industry also depends on the timeliness in which products are delivered to a destination. Promptness is of the utmost importance, as delayed delivery can cause significant losses to the recipient of the consignment. That's why in my opinion it is very interesting to explore how mathematical methods can be applied in logistics and what benefit they can bring.

The aim of this article is to show how optimization and simulation methods can be applied. Application of mathematical methods is based on hypothetical examples because of the absence of real data. In spite of this I think that the obtained results are useful and helpful and achieved conclusions can help in solving a real problem. For more information about operation research methods in logistics please refer to [12], [13], [15].

2 Logistics technologies

At the beginning it is necessary to define what the logistics technologies actually are. According to [11] we can define logistics technologies as methods by means of which we try to select and arrange logistics operations into systems. Their goal is to provide the required level of logistics services at the lowest cost or to achieve the maximum level of logistics services at a given level of costs. This system, which is understood as sequence of processes, operations and treatments, and arranged in steady sub-processes, can be defined as a logistics technology.

The logistic practice applies a wide scale of such systems. These technologies can be divided into several groups: classical, telematic, virtual and complex [11]. The group of complex logistics technologies and systems includes Just in Time (JIT), Kanban, Quick Response (QR), and Efficient Consumer Response (ECR). This work is focused mainly on Just in Time and Kanban technologies.

2.1 Just in Time

According to [6] and [11] JIT is a management approach, which was discovered in Japan in the 1950s. It was subsequently adopted by Toyota and many Japanese manufacturing establishments with considerable success in raising productivity by eliminating waste. JIT is based on the delays' elimination, detection of which is proceeding from production process towards the external transport and satisfying the needs for particular material in the manufacture, by delivering products *Just in Time*. Products are delivering in small quantities, as late as possible. Deliveries are very frequently realized. This principle allows the segments in supply chain to be established with minimum safety reserve. In this case the company becomes fully dependent on its suppliers, and delay from their side can cause the delay of the full production process.For the successful implementation of JIT two basic requirements have to be fulfilled. The first requirement is an important change in the supplier and customer relationships, and the second one is perfect coordination between all participants of the supply chain.

2.2 Kanban

Kanban is a stock-free technology that came from Japan as well as JIT. It is actually one of modifications of JIT principle. Kanban is a Japanese word, where "kan" means "visible" and "ban" means "label", "tag". It is the name of a production transport card, which is provided with an empty transport for visualizing information about deliveries.These shipping notes (transport cards) are also playing role of the controlling documents for production process. They include the following information: name and number of the product, a description (size, weight), identification number of the shipping note and the name of supplier and customer.

The process of Kanban system is going on as follows:

- the arrival of an empty container is the impulse for starting theproduction of the batch,
- the produced batch is put into container according to the information in the transport card and sent to user,
- user takes over the delivered batch and checks characteristics of the delivered items.

For more detailed information about these requirements please see [11], [14].

3 JIT application

3.1 Methods and data

Before starting the analysis of the examples I would like to specify the methods, which were applied, and data used.

I consider a hypothetical factory, which produces components for automobiles. It currently has a reasonable profit and management wants to renovate it and increase the efficiency of whole process of the company running. The company has one plant now and one of the main aims is to expand output. Management is thinking about the construction of two factories and at the same time they want the firm to run under the principle of Just in Time. That means that the technology must be applied not only in the processes of enterprise, but also in the entire logistics chain.

For the successful implementation of JIT a profound change in the relationship between supplier and customer has to be come to pass. Cooperation with existing suppliers is not satisfying enough, that's why managers want to start working with seven new suppliers. They have to make a choice between fifteen potential partners. Below there is a range of criteria that describe candidates:

• Supply price,

• Product quality,

• Wastage rate,

• The level of customer services

• Distance from the supplier company,

• Personal relationships between company staff and employees of the supplier,

• Delivery flexibility,

• Reference from other customers,

• Existenceof certification according to ISO 9001standard,

• Supply punctuality,

• Quality of the vehicle fleet.

Information about all candidates is summarized in the following table.

	supply price [CZK]	quality [points]	wastage rate [%]	customer services [points]	distance [km]	personal relationships [points]	delivery flexibility [points]	other customer references [verbally]	ISO 9001 certification [yes/no]	supply punctuality [points]	quality of the vehicle fleet [verbally]
supplier 1	72	7	5	10	120	10	3	very good	yes	7	very modern
supplier 2	57	6	6	6	53	9	4	excellent	yes	9	mixed
supplier 3	56	3	7	8	15	10	4	excellent	yes	7	new
supplier 4	73	9	2	4	95	9	7	very good	yes	5	out of date
supplier 5	86	9	1	9	46	3	8	sufficient	no	3	out of date
supplier 6	83	6	5	3	36	7	2	good	no	6	new
supplier 7	33	7	5	8	111	7	3	very good	yes	7	very modern
supplier 8	48	7	5	8	103	2	9	sufficient	no	4	mixed
supplier 9	69	7	4	7	67	6	8	very good	yes	3	obsolete
supplier 10	62	4	8	7	80	5	7	good	no	6	very modern
supplier 11	47	1	10	6	52	1	4	sufficient	no	8	obsolete
supplier 12	93	9	0,5	6	30	6	5	good	no	6	obsolete
supplier 13	89	3	8	6	91	8	7	excellent	yes	9	mixed
supplier 14	57	7	4	6	28	8	9	very good	no	8	mixed
supplier 15	53	8	4	4	49	9	5	very good	yes	8	new
criteria type	min	max	min	max	min	max	max	max	max	max	max

Table 1 Input data

To compare the suppliers I used one of methods of multi-criteria evaluation of alternatives.

The model of multi-criteria evaluation of alternatives contains a list of alternatives $A = \{a_1, a_2, \dots, a_p\}$, a list of criteria $F = \{f_1, f_2, \dots, f_k\}$ and an evaluation of the alternatives by each criterion in the criteria matrix [4]:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & \vdots & & \vdots \\ y_{p1} & y_{p2} & \cdots & y_{pk} \end{bmatrix},$$

where y_{ij}, $i = 1, 2, \dots, p$, $j = 1, 2, \dots, k$ represent information about the evaluation of each alternative by each criterion [2], [4], [7].

For the selection of suppliers, I decided to use TOPSIS method. The basic concept of this method is that the best alternative should have the shortest distance from the ideal alternative and the farthest distance from the basal alternative. The method is also able to rank alternatives using the relative index of distance of the variant from the basal variant. Higher relative index of distance means better variant. For solving problems with this method user should have additional information in the representation of weights for each criterion [7].

To proceed in solving this problem we need additional information about the order of importance of each criterion. I decided to use the ranking of the criteria by M.R. Lindere and H.E. Firon [9]. On the opinion of those authors the most important criteria are quality of products and supply punctuality. The way the other criteria were settled in, can be found in the following table. Based on the established order weights of individual criteria were calculated.

	supply price	quality	wastage rate	customer services	distance	personal relationships	delivery flexibility	other customer references	ISO 9001 certification	supply punctuality	quality of the vehicle fleet
ranking	3	1	4	6	7	9	5	8	10	2	11
values	9	11	8	6	5	3	7	4	2	10	1
weights	0,136	0,167	0,121	0,091	0,076	0,045	0,106	0,061	0,030	0,152	0,015

Table 2 Criteria ranking and weights

For further calculation data have to be adjusted: convert all the criteria for maximization type and quantify the verbally entered data.

	supply price	quality	wastage rate	customer services	distance	personal relationships	delivery flexibility	other customer references	ISO 9001 certification	supply punctuality	quality of the vehicle fleet
supplier 1	21	7	5	10	0	10	3	7	1	7	10
supplier 2	36	6	4	6	67	9	4	10	1	9	5
supplier 3	37	3	3	8	105	10	4	10	1	7	8
supplier 4	20	9	8	4	25	9	7	7	1	5	1
supplier 5	7	9	9	9	74	3	8	1	0	3	1
supplier 6	10	6	5	3	84	7	2	4	0	6	8
supplier 7	60	7	5	8	9	7	3	7	1	7	10
supplier 8	45	7	5	8	17	2	9	1	0	4	5
supplier 9	24	7	6	7	53	6	8	7	1	3	3
supplier 10	31	4	2	7	40	5	7	4	0	6	10
supplier 11	46	1	0	6	68	1	4	1	0	8	3
supplier 12	0	9	9,5	6	90	6	5	4	0	6	3
supplier 13	4	3	2	6	29	8	7	10	1	9	5
supplier 14	36	7	6	6	92	8	9	7	0	8	5
supplier 15	40	8	6	4	71	9	5	7	1	8	8
criteria type	max	max	max	max	max	max	max	max	max	max	max

Table 3 Adjusted input data

3.2 Results

For solving this problem I used the Euclidean distance of measure. Based on the values d_i^+, d_i^- relative distance indicator c_i was calculated. Then it was necessary to organize all of variants according to c_i values from the largest to the smallest. I used Excel for solving this model. The task was to choose the best seven suppliers from fifteen candidates. The 14th, 15th, 7th, 2nd, 9th, 1st and 5th supplier were chosen.

3.3 The optimization model of the logistics chain in JIT

Another important assumption of JIT is perfect coordination between all elements of the supply chain. In this example, we assume that the logistics chain consists of seven suppliers, which have been selected in the previous calculation, three producers (factories), two distributors and five potential customers. In JIT applying the main objective is to set the flow between elements of the supply chain that minimizes total costs and balances the production flow. It's also necessary to respect the capacity of participants of the chain and the largeness of customers' demand. I decided to solve this task by model of mixed linear programming. The aim in this case is to find the optimal size of the production flow between participants of the chain, which is minimizing total costs of the whole system.

Due to limited size of article the data used are not introduced here. Author can provide with data on demand. A mathematical model was formulated based on an optimization model of power supply networks [3].

The mathematical model of the problem is

Minimize

$$z = \sum_{j=1}^{3} F_j Y_j + \sum_{k=1}^{2} F_k Y_k + \sum_{i=1}^{7}\sum_{j=1}^{3} c_{ij} x_{ij} + \sum_{j=1}^{3}\sum_{k=1}^{2} c_{jk} x_{jk} + \sum_{k=1}^{2}\sum_{l=1}^{5} c_{kl} x_{kl}$$

Conditions are

$$\sum_{j=1}^{3} x_{ij} \le a_i, \qquad i = 1, 2, \dots, 7,$$

$$\sum_{k=1}^{2} x_{jk} \le b_j Y_j, \qquad j = 1, 2, 3,$$

$$\sum_{l=1}^{5} x_{kl} \le w_k Y_k, \qquad k = 1, 2,$$

$$\sum_{k=1}^{2} x_{kl} = d_l, \qquad l = 1, 2, \dots, 5,$$

$$\sum_{i=1}^{7} x_{ij} - \sum_{k=1}^{2} x_{jk} \ge 0, \qquad j = 1, 2, 3,$$

$$\sum_{j=1}^{3} x_{jk} - \sum_{l=1}^{5} x_{kl} \geq 0, \qquad k = 1,2,$$

$$Y_{j,k} \in \{0;1\},$$

$$x_{ij}, x_{jk}, x_{kl} \geq 0,\ i = 1,2,\dots,7,\ j = 1,2,3,\quad k = 1,2,\ l = 1,2,\dots,5,$$

x_{ij}... production flow between supplier i and producer j,
x_{jk} ... production flow between producer j and distributor k,
x_{kl} ... production flow between distributor k and customer l,
Y_j ... binary variable that reflects the construction of factory j,
Y_k ... binary variable that reflects cooperation with distributor k,
a_i, b_j, w_k … the capacity of supplier i, producer j, distributor k,
d_l ... demand of customer l,
c_{ij}, c_{jk}, c_{kl} ... unit costs for transport between nodes of different levels of the supply chain,
F_j, F_k ... fixed costs for factory j construction and for cooperation with distributor k.

The objective function minimizes the total cost between all levels of the supply chain. Then there are conditions for not exceeding the capacity of suppliers, producers, distributors and complying with customer requirements. The last two conditions ensure the absence of inventory in stocks of all participants.

After solving this model in Lingo program following results were obtained. Producers will work only with the top four suppliers; collaboration with other suppliers would not be effective for the logistics chain. Vector of binary variables Y_j contains all ones; what means that two more factories will be built. Vector of variables Y_k is unitary as well, what means that producers will work with both distributors. Then distribution centers will deliver goods to all five customers. Minimum total costs for transportation between all participants of the logistics chain, for two new factories construction and for cooperation with distributors are equal to 96 810 000 CZK.

In conclusion I can add that on my point of view linear programming is a suitable tool for the logistics processes modeling. But it also has its cons. The first disadvantage is that optimization model refers to the whole logistics system, what can't guarantee that the recommended changes will bring benefit for each participant. To ensure minimum costs for all levels of the system continuous multi-criteria decision models can be used. The second problem is obtaining the necessary data to create the model. To create a model, which captures all aspects of reality, the analyst must have all necessary data from all entities that is not so simple to achieve.

4 Kanban application

I consider a hypothetical example where there is a manufacturing guild with four machines. Kanban technology was already implemented there, that means machines are moving components using kanban transport cards. The production is set in the way that the assembling can begin either on the first or on the second machine. Then process can proceed to the third and the fourth machine. To analyze the operation of the system I used simulation model where entities are components running through the production process. The aim is to analyze the use of resources and also analysis of emerging queue.

Simulation is method that uses a computer model of a real process to enable managers predicting the behavior of the system, optimizing business processes with respect to criteria and comparing alternatives of different process organizations [1]. The basic idea of the simulation is an imitation of the real business process and monitoring of the consequences of possible changes. This model is based on discrete-event simulation method. Discrete model considers that changes can be caught only at discrete points in time, not continuously [1].

To create a simulation model, I used the program SIMUL8, which is a software package for Discrete Event Simulation (www.simul8.com). Model in SUMUL8 is shown on the picture below.

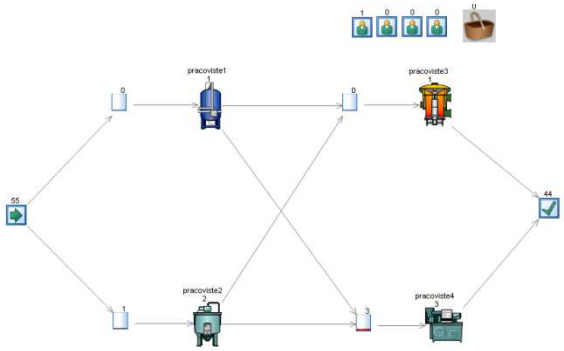


Figure 1 Model of manufacturing guild in SIMUL8

The process begins with the entrance of entities into the system. Subsequently, the first part of entities goes to the first work center and the second part to the second work center. Afterwards entities may proceed to either the third or the fourth work place. Each workplace is operated by resource, which is worker in this case. Kanban container is also defined in the model as a resource. At the beginning I set up the number of containers on the number 3. In this case kanban containers are used on more than 95%, and there are relatively large queues between each work center. The problem of large queues can be solved by increasing the number of moving kanban containers. After increasing the amount of containers on seven, queues arising between the machines are not as long as before. Results received for the first work center are shown on the figure below.

Queue for pracoviste1		
	Minimum queue size	0.00
	Average queue size	0.24
	Maximum queue size	2.00
	Minimum Queuing Time	0.00
	Minimum (non-zero) Queuing Time	0.04
	Average Queuing Time	4.82
	Average (non-zero) Queuing Time	8.90
	Maximum Queuing Time	21.82

Figure 2 Queue characteristics in case of seven containers for the first work center

This example is of course a considerable simplification of the reality and the analysis of real situation using the simulation model will be more complicated. Nevertheless, in my opinion simulation is a suitable tool for modeling the production system in the case of applying Kanban technology.

5 Conclusions

The aim of my article was application of optimization and simulation methods to practical examples in conjunction with logistics technologies. After all the calculations and analyzes I can summarize that the main disadvantages of mathematical models are necessity of taking into account all conditions and possible unavailability of all necessary data. But still I think that the optimization and simulation methods can greatly help in streamlining logistics processes.

References

[1] Dlouhý, M., Fábry, J., Kuncová, M., Hladík, T.: *Simulace podnikových procesů*. Brno: Computer Press, 2007. ISBN 978-80-251-1649-4.

[2] Evans, G.W.: An Overview of Techniques for Solving Multiobjective Mathematical Programs. *Management Science*. 1984, vol. 30, No 11.

[3] Fiala, P.: *Dynamické dodavatelské sítě*. Praha: Professional Publishing, 2009. ISBN 978-80-7431-023-2.

[4] Fiala, P.: *Modely a metody rozhodování*. Praha: Oeconomica, 2006. ISBN 80-245-0622-X.

[5] Janakiraman, S.: *Importance of logistics industry for growing economies*. Seoul: Asia-Pacific Report, 2011.

[6] Kee-hung L., Cheng, T.C.: *Just-in-Time Logistics*. Farnham, Surrey: Ashgate Publishing, 2010.

[7] Kuncová, M., Doucek, P.: Comparison of the cluster analysis and the methods of the multi-criteria evaluation of alternatives used to create groups of countries similar in the take up of the internet services. In: *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011*. Prague, 2011.

[8] Lamber, D., Stock, J. R., Ellram, L.: *Logistika – příkladové studie, řízení zásob, přeprava a skladování, balení zboží*. Brno: CP Books, a.s., 2005. ISBN: 80-251-0504-0.

[9] Lindere, M.R, Firon, H.E.: *Purchasing management and stocks. Logistic / the Lane with English*. Saint-Petersburg: Victoria plus, 2002.

[10] Pastijin H. Leysen J.: *Constructing an Outranking Relation with ORESTE*. Mathematical and Computer Modeling, Vol.12, No 10/11, 1989.

[11] Pernica, P.: *Logistika (Supply chain management) pro 21. století*. Radix spol. s.r.o., Prague, 2004.

[12] Rajesh, R., Ganesh, K., Lenny Koh, S.: *Logistics Design and Modeling - A Simulation Perspective*. 2011.

[13] Rossetti, M., Pohl, E., Limp, F.: *Applications of GIS and Operations Research Logistics Planning Methods for Arkansas Rural Transportation Emergency Planning*. 2008, s. 129.

[14] Sixta, J., Mačát, V.: *Logistika – teorie a praxe*. Brno: CP Books, a.s., 2005. ISBN 80-251-0573-3

[15] Takai, E.: *The Role of Operations Research towards Advanced Logistics*. Science & Technology Trends. 2008, s. 120.

Models of coordination games

Petr Fiala[1]

Abstract. Coordination games attract many theoretically and experimentally oriented economists. In coordination games equilibria in pure strategies are characterized by the requirement that players choose the same strategy. Players are embedded in a graph that represents the pattern of partnerships.
The cooperative approach can be considered as a natural development of the theory that studies the formation of coalitions. This theory analyzes the external structure of coalitions by focusing on the allocation of players within the various groups. Non-cooperative approaches are used for analyzing strategic effects of local interaction for a coordination game with agents exogenously located on network nodes.
The assumptions of basic concepts can be extended to get more realistic models of coordination problems. The approaches for analyzing the coordination problems can be modified also. Possible extensions of basic models given by indirect links, system dynamics, costs and benefits, and imperfect reliability are studied in the paper. Combined approaches for analyzing coordination problems are much promised because they exploit advantages from both theories, cooperative and non-cooperative ones.

Keywords: game theory, coordination, graphs, cooperation, non-cooperation.

JEL Classification: C70
AMS Classification: 91A43

1 Introduction

Coordination is the process of managing interdependencies between activities (Demange and Wooders, 2005). An agent is a subject that has some knowledge about the situation and can perform actions. The problem of coordination occurs when an agent has a choice in its actions, and the choice, the order and the time at which actions are carried out affects performance. A multiagent system is a collection of agents that coexist in an environment, interact with each other, and try to optimize a performance measure. Research in multiagent systems aims at providing principles for the construction of complex systems containing multiple independent agents and focuses on behavior issues in such systems. A key aspect in such systems is the problem of coordination: the process that ensures that the individual decisions of the agents result in jointly optimal decisions for the group.

Networks are presented as environments in which strategic interaction takes place. The networks' topologies and the characteristics of links influence the strategies of all agents and economic results consequently. Networks play a crucial role in the formation of economic structures, in the circulation of information and in the emergence of competition and cooperation among agents.

In principle game theoretic techniques can be applied to solve the coordination problem. Whenever agents should choose an identical action to receive better performance, we speak about a coordination game. As a typical example think about competing technology standards. When agents are capable of agreeing on the technology to apply, the performance is high. However, in the lack of a common technology standard the performance is lower. Game theory provides a collection of tools for predicting outcomes for a group of interacting units. It is the basis for development of network coordination models, cooperative and non-cooperative models as well. We will use normal form games and also extensive form games for analyzing cooperative and non-cooperative behavior of units in the network. In normal form games all units choose their actions simultaneously, whereas in extensive form games units may choose their actions in different time periods. There are some special types of games to illustrate network economy problems as coordination games. A coordination game is a situation with interdependent decisions, a coincidence of interests, and at least two proper coordination equilibria. Nash equilibrium concept is used for solution of non-cooperative models. A basic definition is this: If there is a set of strategies for a game with the property that no player can benefit by changing his strategy while the other players keep their strategies unchanged, then that set of strategies and the corresponding payoffs constitute Nash equilibrium.

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech Republic, pfiala@vse.cz

2 Basic model of coordination game

Coordination games attract many theoretically and experimentally oriented economists. The basic model is a simple symmetric normal form 2 × 2 games which are characterized by having two equilibria in pure strategies (Bala and Goyal, 2000). In coordination games equilibria in pure strategies are characterized by the requirement that players choose the same strategy. Both symmetric equilibria may be candidates for strategy selection.

The game is characterized by the payoff table (Table 1)

	X	*Y*
X	b,b	c,d
Y	d,c	a,a

Table 1 Payoff table

with $a > b > c > d > 0$ and $(b - d) > (a - c)$, i.e. (Y, Y) is Pareto dominant equilibrium and (X,X) is the risk dominant equilibrium.

Let $N = \{1, 2, \ldots, n\}$ be the set of agents. Agents play an n-player game. Agents are embedded in a graph G that represents the pattern of partnerships. The link between i and j is denoted by g_{ij} such that $g_{ij} = 1$ when i and j are partners and $g_{ij} = 0$ otherwise. The game is played in large populations with players who are pairwise matched in network structures. Players are allowed to choose their neighbors by themselves and to select simultaneously their neighbors in the population and the strategy in the basic game.

Decision making is supposed to be deterministic. Opening a new link to a member of the population is supposed to generate constant connection costs per link for the agent who initiates the link. Let k denote the linking costs. The relative size of linking costs compared with the payoffs of the game has a big impact on the resulting equilibrium network in the population. The resulting equilibrium networks are characterized by non-directed graphs.

Depending on the particular value of linking costs k we obtain different graphs for coordination games:

a) If $k > a$ then the unique equilibrium network G is the empty network and the action choice of each player in the coordination game is not determined.

b) If $k < d$ holds then the unique equilibrium network G is the complete graph and all players choose the equilibrium strategy X or all players choose the equilibrium strategy Y.

c) If $d < k < c$ holds then either the equilibrium is the one obtained in part *b)* or there is an equilibrium such that the resulting network is characterized by a graph G with two sets of vertices N_1 (agents play the strategy X) and N_2 (agents play the strategy Y) such that all vertices in N_1 are connected with all vertices in N_2 but not with each other in N_1 while all vertices in N_2 are also connected with each other.

d) If the relation $c < k < b$ holds then either the equilibrium is the one obtained in part *b)* or an equilibrium induces a disconnected graph G with two components where each component is a complete graph and players in one component (N_1) choose action X, players in the other component (N_2) choose action Y.

e) If $b < k < a$ holds then an equilibrium network G is the complete graph with all players choosing Y. An alternative equilibrium is the empty graph with all players choosing X.

Figure 1 Typical equilibrium networks

3 Solution approaches

The approaches for analyzing the coordination problem and network formation process can be divided to two groups: cooperative and non-cooperative approaches.

3.1 Cooperative approaches

An important first paper in the literature is by Myerson (1977). Myerson started from cooperative game theory. The model is a game $(N; v)$, where N is a set of players and v is a characteristic function denoting the worth $v(S)$ of each coalition S. He defines a cooperation structure as a non-directed graph G among the agents. A graph represents a list of which individuals are linked to each other, with the interpretation that if individual i is linked to individual j in the graph, then they can communicate and function together.

Thus, a network G partitions the set of individuals into different groups who are connected to each other. The value of a coalition S under the network G is the sum of the values of the sub-coalitions of S across the partition of S induced by G.

The cooperative approach, introduced by Myerson (1977), can be considered as a natural development of the theory that studies the formation of coalitions: this theory analyzes the external structure of coalitions by focusing on the allocation of players within the various groups only; the purely cooperative approach goes further and considers the set of communication components within each coalition.

Myerson's approach was followed by Aumann and Myerson (1988) who introduced two-stage game: in the first stage links are formed, and in the second stage the players receive payoffs which depend on the value of the network, according to an exogenous rule. They use the extensive form to represent the game, assume that players don't bear any cost while forming links.

Jackson and Wolinsky (1996) introduce the concept of pairwise stability in order to analyze the network formation process. The game is given by the finite set of players N, function $v : G \rightarrow R$ gives the value of a network G, while $\pi(G, v)$ is the payoff function describing the way in which the value of the network is allocated among players. A network G is pairwise stable if, given any two linked agents, none of them benefit from deleting the link, and if, given any two unlinked agents, it cannot be that both of them find it convenient to form the link.

Let G be a network; $(i, j) \in G$ means that players i and j are linked to each other, $(i, j) \notin G$, means the opposite, $G + (i, j)$ is the network obtained when the link between i and j is set and $G - (i, j)$ means just the opposite.

A network G is pairwise stable if both of the following conditions hold

$$(i)\ \forall (i, j) \in G, \pi_i(G, v) \geq \pi_i(G - (i, j), v), \pi_j(G, v) \geq \pi_j(G - (i, j), v) \quad (1)$$

$$(ii)\ \forall (i, j) \notin G, \text{ if } \pi_i(G + (i, j), v) \geq \pi_i(G, v), \text{ then } \pi_j(G + (i, j), v) < \pi_j(G, v) \quad (2)$$

The concept of pairwise stability is very helpful to both evaluate a given network topology and determine the set of equilibria, but it says nothing about their desirability. Pairwise stable networks are not always efficient.

3.2 Non-cooperative approaches

Bala and Goyal (2000) studied the strategic effects of local interaction for a coordination game with agents exogenously located on network nodes. There are some assumptions of the approach. Players can add or delete links unilaterally; thus there is no need for an agreement between the players in order to a link to be set. Consider two different patterns of information transmission: in the one way flow models, the information goes from j to i, while in the two-way flow models, the information goes from and to both agents.

Nash Network is based on Nash equilibrium concept for the network G:

- In the one-way flow models, a Nash Network is either empty or minimally connected.
- In the two-way flow models, a Nash Network is either empty or minimally pairwise-connected.

Bala and Goyal show that, in general, the number of Nash networks increases very rapidly with n. That leads to focus on strict Nash networks that are networks supported by a strict Nash equilibrium. The set of strict Nash networks is significantly more restrictive:

- In the one-way flow models, a strict Nash network is either empty or a wheel.
- In the two-way flow models, a strict Nash network is either empty or a center- initiated star.

Figure 2 Typical strict Nash networks

4 Extensions of basic concepts

The assumptions of basic concepts can be extended to get more realistic models of coordination problems. The approaches for analyzing the coordination problems can be modified also.

We mention only some of these possible extensions:

- indirect links,
- system dynamics,
- costs and benefits,
- imperfect reliability.

4.1 Indirect links

In the model, agents are information seekers who gain utility from having more information. A pair of agents can play the game even if connected only through indirect links. The cost of information $h(v_j, d_{ij})$ depends on the value of the information itself and the distance it traverses in the network. When the costs of information increase with distance, the complete network is the only Nash network. When cost of information decreases with distance, every Nash network is either a tree or is empty. However, not every tree is Nash.

There are some possibilities how to model dependency of value of agent *j*'s information to *i* on distance between *i* and *j*. The concept of information decay introduces the decay parameter to modify the value of information by distance between agents. The paper analyzes information flow in the absence and presence of information decay.

The payoff to player *i* from the network *G* without double links is given by

$$\pi_i(G) = \sum_{j \in G(i)} (v_j - h(v_j, d_{ij})) \tag{3}$$

For analysis we construct easer payoff function by assuming information of higher value is more expensive and information coming through shorter paths in the network is more expensive. Information that comes through longer paths involves a longer waiting time and hence is cheaper.

$$\pi_i(G) = \sum_{j \in G(i)} (v_j - \frac{v_j}{d_{ij}}) \tag{4}$$

We will study the implications of information decay, i.e., the value of information decreases as it traverses through the network. This will create an incentive for the agents to form shorter paths.
Assuming Information that comes through longer paths in the network is more expensive, information decay reinforces every agent's incentive to establish a direct link to all the other agents leading to the complete network.

Assuming that information that comes through shorter paths in the network is more expensive. To study the impact of information decay a variation of the payoff function in given form by equation (4) is introduced.

Let

$$\pi_i(G) = \sum_{j \in G(i)} (v_j(1 - \delta d_{ij}) - \frac{v_j}{d_{ij}}) \tag{5}$$

where $\delta \in (0,1)$ is the decay parameter. A direct link now yields a payoff of $-\delta v_j$ and as before this form enables us to focus on the indirect links.

There is possible to analyze relations between a value of decay parameter and Nash network architecture. We define a network structure. A periphery-initiated star is the acyclic network where all the other (*n*-1) agents initiate exactly one link with the central agent. Then it is possible to formulate and to prove next theorem.

Let the payoffs be given by formula (5). Then the periphery-initiated star with the central agent having minimum value is Nash for

$$1/6 < \delta \leq 1/4. \tag{6}$$

4.2 System dynamics

Problems of coordination on networks are dynamic problems very often. For example network externalities as positive or negative are dynamic effects. The situation can be described by standard instruments like differential equations. The network structure of the problems introduces the possibility of using extensive form games and repeated games for describing the dynamic processes.

The dynamics of network formation has powerful effects on individual behavior. If costs forming links are below a certain threshold then agents coordinate on the risk dominant action, while if costs are above this threshold then they coordinate on Pateto dominant action.

A simple dynamic approach is possible to use. Assuming discrete time $t = 1, 2, \ldots$, at every time period t the state of the system is given by the strategy profile S(t) specifying action, and links established, by each agent. At every time period t, there is independent probability p that any agent gets a chance to revise his strategy. If he receives this opportunity, he selects a new strategy

$$s_i(t) = \underset{s_i \in S_i}{\operatorname{argmax}} \pi_i(s_i, s_{-i}(t-1)). \tag{7}$$

He selects a best response to what other agents chose in the preceding time period.

The strategy revision process defines a Markov chain on S. There are several absorbing states of the Markov chain that means multiple strict equilibria. This motivates the examination of the relative robustness of each of them.

4.3 Costs and benefits

Standard models can be modified by introducing cost and benefit structures according the real situation. The situation can differ in assumptions who will share link formation costs and related benefits and what are rules of sharing.

Individual efforts to balance the costs and benefits of forming links generate, over time, simple network architectures. Equilibrium networks are either complete or have the star architecture. The nature of individual behavior in the coordination games is sensitive to the cost of forming links and this relationship is robust across the settings. Agents choose the risk dominant action for low costs of forming links but the Pareto dominant action for high costs.

There is the relationship between costs and individual behavior in coordination games. When the cost forming links is small, agents wish to be linked with everyone irrespective of the actions they choose. From individual perspective, the relative attractiveness of different actions is quite insensitive to the network structure and the agent will operate in a complete network. The risk dominant and inefficient convention prevails since this convention is harder de-stabilize than Pareto dominant but risk dominated one.

If costs of forming links are high, individual agents choose to form links only with those who play similar actions in the coordination game. But then, since the payoff from coordinating on Pareto dominant action is higher, the number of players required to de-stabilize the inefficient convention must also be higher than for the efficient one and the latter convention therefore prevails. The particular relationship between the costs of forming links and the nature of coordination is linked to how strong are the incentives of players (in particular, those playing Pareto dominant action) to separate themselves from those playing a different action.

4.4 Imperfect reliability

In order to represent the assumption of imperfect reliability and to add realism there is possible to construct a stochastic network model. By introducing a probability parameter p into stochastic network model each link works with probability p, while fails with probability 1-p.

5 Conclusion

Network coordination is the important subject of an intensive economic research. The paper presents a basic theoretical framework for analyses of network coordination and some problems and approaches for coordination activities. The framework makes possible to develop simple models for analyzing specific features in network coordination.

Extensions of the basic framework are possible and very useful to be more realistic. Than it can be construct simple models for analyzing coordination problems. The simple models have very important managerial implications indeed. The combination of such models can give more complex views on the problem of network coordination.

The approaches for analyzing the coordination problem and network formation process can be divided to two groups: cooperative and non-cooperative approaches. Combined approaches for analyzing coordination problems are much promised because they exploit advantages from both approaches.

Acknowledgements

Supported by the grant No. 13-07350S of the Grant Agency of the Czech Republic and by Grant No. IGA F4/19/2013. Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Aumann, R.J., Myerson, R.B. (1988): Endogenous Formation of Links between Players and of Coalitions: An Application of the Shapley Value. In *The Shapley Value*, A. E. Roth (ed.), Cambridge: Cambridge University Press, 175-191.

[2] Bala, V., Goyal, S.: A Noncooperative Model of Network Formation. *Econometrica*, vol. 68 (2000), 1181-1229.

[3] Demange, G., Wooders, M. (eds): *Group Formation in Economics: Networks, Clubs and Coalitions*. Cambridge University Press, Cambridge, 2005.

[4] Jackson, M., Wolinsky, A.: "A Strategic Model of Economic and Social Networks. *Journal of Economic Theory*, vol. 71 (1996), 44-74.

[5] Myerson, R.: Graph and cooperation in games. *Mathematics of Operation Research*, vol. 2 (1977), 225-229.

Robustness in estimated macroeconomic policy factors for the Czech Republic

Tomáš Formánek[1], Roman Hušek[2]

Abstract. Macroeconomic policy tools' selection should be based on robust quantitative analyses. Given the unstable global environment, actual macroeconomic policy actions should be based on relevant outputs from competing modelling paradigms, as contrary to choosing one type of optimization methodology, despite of its arguable sophistication and simulation performance. Additional insight and increased prudency may be achieved by comparing the pre-crisis and post-crisis economic dynamics for diverse policy factors in different small open economies (SOEs) in order to quantify the effects and statistical properties of discretionary policy tools available to the central authorities, both monetary and fiscal.
In this contribution, we use alternative quantitative means to evaluate the robustness of macroeconomic policy evaluation factors for the Czech Republic and other SOEs. Both observable and unobservable variables are considered and VAR models with alternatively specified impulse response functions are used.

Keywords: macroeconomic policy factors, robust estimates, vector autoregression.

JEL Classification: C18, C32, E61
AMS Classification: 91B55

1 Introduction

As the direct negative impacts of the 2008 crisis diminish, we may be able to use econometric techniques to assess and quantify the nature of its legacy using the available post crisis macroeconomic data observations that have accumulated and are available for analysis. Alternatively specified vector autoregression (VAR) models with diverse impulse response function (IRF) identification methods, state-space models, static and dynamic panel data-based models and other types of models as based on the pre-crisis and post-crisis data may allow us to evaluate macroeconomic dynamics and relationships for both observable and unobservable macroeconomic variables. Our approach may also be used to address the issue of robustness (stability) of the estimated dependencies between variables as well as the robustness of predictions and IRFs generated from different types of model specifications. Given the limited amount of data observations available for single economy-based models, we explore the application of selected panel data estimation techniques. However, strong theoretical and practical limitations apply for inclusion of lagged endogenous variables into panel data models.

This paper is organized as follows: Next section provides both theoretical background for our research and references to additional literature resources. Chapter three describes the data and data transformation process used in this paper. In chapter four, we provide key empirical results along with corresponding discussion and interpretation. Last section and the list of references conclude our contribution.

2 Literature review

Applied macroeconomic analysis often involves setting up and estimating unrestricted VAR models that are used to provide information about the responses of economic variables to diverse shocks. However, in order to calculate the IRFs of interest, additional transformation to the estimated model is necessary in most cases, as innovations from different VAR model equations are usually correlated. For an m-variable (m-dimensional) VAR model, such contemporaneously correlated innovations $\boldsymbol{u}_t' = (u_{1t}, u_{2t}, \ldots, u_{mt},)$ may be interpreted as having a common component that cannot be attributed to an individual endogenous variable of the VAR model. Fortunately, many well established transformation methods are available and built into most econometric software packages. Despite of their ease of use, we should bear in mind that virtually all transformation methods necessitate prior identification assumptions (i.e. restrictions) which themselves may not be subject to statistical verification.

IRFs resulting from the so called **Cholesky orthogonalization** of shocks as described e.g. in [9] depend on the choice of ordering of variables. Recursive (one-way) contemporaneous dependencies imposed by this method attribute the whole effect of the common component in innovations to the variable that comes first in the

[1] VŠE Praha, nam. W. Churchilla 4 Praha 3 Prague 130 67 Czech Republic. e-mail: formanek@vse.cz
[2] VŠE Praha, nam. W. Churchilla 4 Praha 3 Prague 130 67 Czech Republic. e-mail: husek@vse.cz

VAR model. Hence, shocks to the first variable affect all endogenous variables, shocks to the second variable influence all variables except the first one, etc. Cholesky-identified IRFs may change significantly as a result of re-ordering variables in the VAR model, unless the contemporaneous correlations in $\boldsymbol{u}_t$ are negligible, which in turn would eliminate the necessity for *any* orthogonalization. **Generalized IRFs** as described in [11] provide good solution to the ordering-dependence problem. However, this is only achieved by generating *variable specific* Cholesky-type IRF factorizations for each *j*-th variable in the VAR model, where the *j*-th variable comes as first in the Cholesky ordering context. A popular approach belonging to the *structural decomposition* class of methods is the **Blanchard Quah decomposition** (BQ). This approach was pioneered in [3] where theoretically based long-term zero value identifying restrictions on accumulated IRFs are imposed in order to decompose the innovations to observable macroeconomic variables into directly unobservable but economically interpretable and conveniently defined shocks such as supply and demand shocks, monetary policy or technological shocks, etc. Unfortunately, this type of identification scheme is applicable only upon meeting relatively strict conditions and subject to availability and validity of theoretically justifiable long term restrictions to be imposed on IRFs calculated from the VAR model (for a given data set and scope of analysis). Model specification, estimation and potential caveats for this type of orthogonalization are discussed in [6].

While VAR models and IRFs are very useful and popular tools for econometric research, they also have some cutbacks that should not be overlooked. Different authors, such as [9] provide detailed discussion on the inherent limitations specific to all VAR models, regardless of their actual specification or estimation method used. Useful theoretical insight into the potential and limits of VAR model specification and IRF interpretation is provided in [5], along with references to additional key articles. In practical terms, IRF interpretation may often be blurred by the fact that most econometric studies are based on low-dimensional VAR models. This implicates a high potential for incompleteness and **omitted variable bias**. VAR-based empirical studies usually do not explicitly address this topic; yet leaving important variables out of the VAR /or vector error correction (VEC) / model describing the data generating process may result in distorted IRFs with possibly misleading interpretations. A detailed discussion is provided in [9].

A feasible approach to deal with potentially misspecified *stationary* VAR models is provided by Jorda [8] who shows a method of IRF calculation through the so called **local projections**. This method is arguably more robust to model misspecification than common (analytically calculated) IRFs. The reasoning behind local projections as in [8] may be summarized as follows: IRFs are forecast functions calculated at increasingly distant horizons and therefore misspecification errors from the underlying VAR model generally do add up with the growing forecast horizon. On the other hand, local projections are a set of short (e.g. one period ahead) forecasts generated at each forecast horizon. Individual forecast horizons are based on sequential regressions of the endogenous variable that is repeatedly shifted multiple time periods (steps) ahead. Robustness of local projection IRFs to model misspecification comes at a price, though. Local projections generate IRFs that are less accurate (i.e. less efficient) than common IRFs. Using a simple model setup, Teulings and Zubanov [13] introduced a formula for such efficiency loss. Along with other authors referenced in [13], they provide improved estimators for local projection IRFs with extensions towards non-stationary VARs and single equation models.

Macroeconomic panel data and relevant estimation methods such as fixed effect (FE) and random effect (RE) estimators allow us to combine information from observations collected both across time and across individual economies. This can improve the quality and significance of the individual regression coefficients and the model as a whole. Moreover, better insight into the macroeconomic dynamics of interest may be achieved. However, we must be very careful when combining panel data methods with lagged variables used as regressors. Such models are usually referred to as **dynamic panel data models** and under most circumstances, they suffer from a **Nickell bias**, as derived in [10]. An intuitive summary of this bias would be the following: FEs take up some part of the dynamic effect and therefore dynamic panel data models lead to overestimated FEs and underestimated dynamic interactions. Whether the Nickel bias is significant in a particular model/dataset situation is an empirical question. Nevertheless, in theory this bias persists unless the number of time observations goes to infinity. The inclusion of additional cross-sections to the dataset would worsen the bias in most cases. Fortunately, the **Arellano-Bond** (AB) **estimator** was introduced in [1] for panel data single equation models. The idea behind AB estimator is straight-forward: the model is transformed into first differences to eliminate the individual effects; then a generalized method of moments (GMM) approach is used to produce *asymptotically efficient estimates for the dynamic coefficients*. The AB estimator is relatively robust and has gained a lot of popularity in recent years. Also, extensions of the AB estimator for VAR models have been developed, as well as other generalizations described in [14]. However, there is one important issue that needs to be observed: *autocorrelation in residuals* of the estimated model renders the AB estimator inconsistent. Interestingly enough, Wansbeek [14] provides a short meta-analysis of 51 papers containing the AB estimator, which were quasi-randomly selected from the Web of Science (i.e. papers subject to refereeing process) for further inspection. In 20 out of those 51 papers, there was no information about autocorrelation test results. From the remaining 31 papers, 7 did reject the null hypothesis of no autocorrelation without further actions being taken.

State space models belong to a very general and flexible class of methods that allow for combining the observable data with potentially unobservable (theoretically derived) prior information. Structural time series models allow us to view observable data series as a sum of unobservable components. For example, observable real GDP series may be dissected into a stochastic trend component (representing the potential product and its dynamics), cyclical (transitory) component and some white noise component. **Kalman filtering**, as thoroughly described in [9], is a specific recursive method for estimating the unobservable *states* and the coefficients describing the dynamics of their transition. Such estimates are based on the *signal* (observable) data and prior information concerning *data generating mechanisms*. Reverting to our real GDP example, we might assume the potential output to follow a random walk with a drift, the cyclical component to be defined through a stationary ARMA(*p*,*q*) process and we would usually expect the random elements to be Normal and i.i.d. Abundant theoretical and empirical resources for state space models are referenced to in [9].

During the post-2008 crisis period, increased requirements are being put on the robustness of economic analyses. In this context, Wieland [15] and many others argue that actual (practical) macroeconomic policy actions should be open to and even based on outputs from competing modeling methods and paradigms. By exploring two or three competing theoretical bases or estimation procedures instead of one, we do not eliminate the model selection bias per se. Yet, such approach would usually lead to increased robustness in economic analysis results and subsequent actions, as contrary to choosing one type of optimization and modeling, despite of the potential sophistication and arguably remarkable simulation performance achieved through any individual model or approach. For example, Wieland [15] provides a review of findings regarding the effectiveness of temporary fiscal stimulus measures from a range of models and studies.

3 Data and their transformation

In order to ensure consistency, all observed data series originate from the International Monetary Fund database (http://elibrary-data.imf.org/). Quarterly series (2000:Q1 – 2013:Q4) are used for all calculations, unless stated otherwise in the text. Real GDP and GDP deflator base indexes were used for analysis and model construction, along with money market interest rates. Data for Czech Republic were combined with observations from Austria, Poland, Germany and Slovak Republic in order to verify and/or improve robustness of the results. We have used EViews software for estimation, graph rendering and most of the data handling.

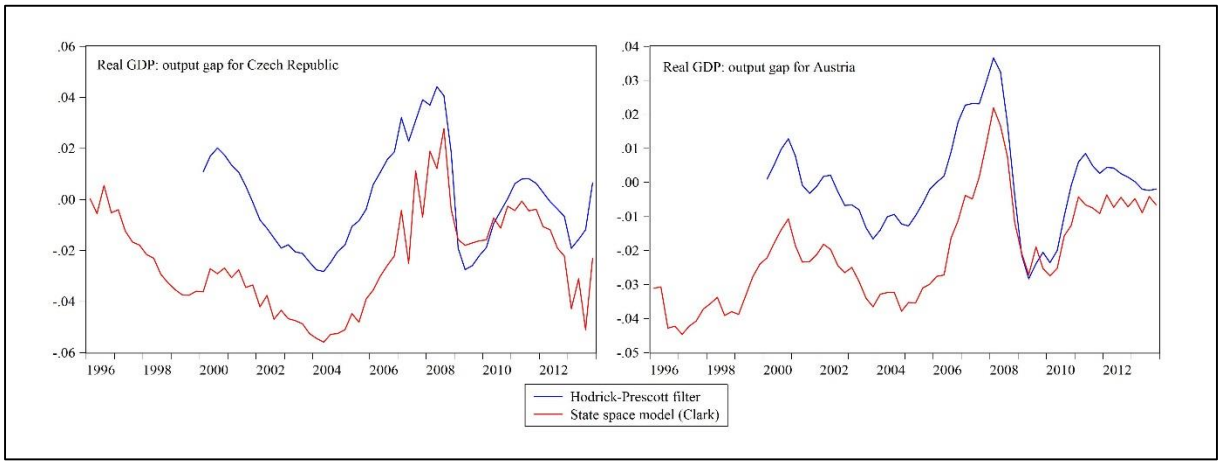


Figure 1 Alternative output gap quantifications for Czech Republic and Austria

For applications presented in this paper, some of the data may not be used in their observed forms; for the amended Shapiro-Watson model (4) - (5), we need to provide time series representing deviations from steady states. Usually, variables such as output gap (defined here as the relative deviation of actual output from its potential value) may be easily calculated using the formula $y_{gap,t} = (y_t - y_t^*)/y_t^*$ where the potential product y_t^* is obtained from the Hodrick-Prescott (HP) trend component as in [7]. For the sake of theoretical and methodological robustness evaluation of the output gap calculation method, we have decided to add a second (competing) approach, derived from the state space model used by Clark [4] (state space model as introduced in [2] might also be applied for this purpose, we omit this approach due to space constraints):

$$Y_t = T_t + C_t \tag{1}$$

$$T_t = \alpha + T_{t-1} + \varepsilon_{1t} \quad , \quad \varepsilon_{1t} \sim N(0, \sigma_1^2), \text{ i.i.d.} \tag{2}$$

$$\phi(\mathrm{L})C_t = \theta(\mathrm{L})\varepsilon_{2t} \quad , \quad \varepsilon_{2t} \sim N(0, \sigma_2^2), \text{ i.i.d.} \tag{3}$$

In the model (1) – (3), the observed (signal) variable Y_t (in logs of real GDP) is made up by two unobservable components: trend T_t follows a random walk with a drift and the cyclical (transient) component C_t is a stationary ARMA(p,q) process. Additionally, a more realistic non-restrictive (diffuse) prior is used for covariance between the errors: $cov(\varepsilon_{1t}, \varepsilon_{2t}) = c_{12}$, instead of the $c_{12} = 0$ default setting for state-space random errors in EViews.

For Czech Republic and Austria, figure 1 may be used to compare the HP-filter based output gaps $y_{gap,t}$ with alternative output gaps based on the approximation: $T_t = y_t^*$ as per model (1) to (3). In order to calculate the Clark (state space) potential outputs and output gaps, we used ARMA(2,1) for C_t and we took advantage of the available Y_t time series (starting in 1996Q1) for both economies. The similarity of resulting output gaps in figure 1 cannot be described as perfect, yet it seems satisfactory. Upon inspecting the overall series dynamics, we may conclude that both methods lead to fairly similar results and all the main turning points are well matched for both methods used and economies shown. Therefore, we would interpret the results as shown in figure 1 as providing solid evidence for the robustness of the HP-filter based output gap estimates. Using alternative ARMA(p,q) settings for C_t, similar evidence for robustness of HP-filter based $y_{gap,t}$ estimates for other EU economies may be provided, as well as for the HP-filter generated inflation gaps used in model (4) – (5). Raw observable data and processed data sets are available from the authors upon request.

4 Empirical results and interpretation

The predicted macroeconomic responses to alternative monetary and fiscal policy scenarios under consideration by the central authorities stand out as key policy-making factors to be considered and evaluated. We use a generalized Shapiro and Watson [12] VAR(2) specification, along with a set of long-term identifying restrictions in order to estimate the expected dynamics of the Czech economy's responses to theoretically defined and conveniently stratified shocks that may be associated with different types of shocks and economic policy actions. In matrix form, our baseline VAR model may be described as follows:

$$\boldsymbol{y}_t = \boldsymbol{\alpha}_t + \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-2} + \boldsymbol{u}_t \,, \tag{4}$$

$$\boldsymbol{\varepsilon}_t = \boldsymbol{G}^{-1}\boldsymbol{u}_t \,. \tag{5}$$

In this model, $\boldsymbol{y}_t$ is a 3x1 vector containing endogenous variables (real GDP, GDP deflator, real interest rate). The first two variables are expressed in terms of gaps (based on HP-filter, using the methodology described in chapter three) and real interest rate is calculated as the difference between money market interest rate and the inflation (based on GDP deflator information available ex-post). All three data series are stationary (ADF tests were applied). $\boldsymbol{\alpha}_t$ is a 3x1 vector of intercepts. Coefficients of the 3x3 $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ matrices define the estimated VAR(2) dynamics. $\boldsymbol{u}_t$ is a 3x1 vector of *random elements*, generated from the VAR model (its elements are potentially contemporaneously correlated). $\boldsymbol{\varepsilon}_t$ is a 3x1 vector of *unobservable and orthogonal random elements*: (supply shock ε_t^{sup}, money market-related demand shock ε_t^{LM}, goods and services market demand shock ε_t^{IS}). $\boldsymbol{G}$ is a 3x3 decomposition matrix. The expression $E(\boldsymbol{u}_t\boldsymbol{u}_t') = \boldsymbol{G}\boldsymbol{G}'$ provides only 3(3+1)/2 = 6 standard (Cholesky-type) identifying restrictions for the 9-element matrix $\boldsymbol{G}$ and its identification is achieved using three additional long term zero value identifying restrictions (BQ-type restrictions) that may be verbally summarized as follows: ε_t^{LM} and ε_t^{IS} demand shocks have no long term effect on real GDP, ε_t^{IS} has no long term effect on GDP deflator. The methodology we may use to produce the matrix $\boldsymbol{G}$, vector $\boldsymbol{\varepsilon}_t$ as well as BQ-type IRFs (responses to individual unobservable shocks) is thoroughly described in [9], while [6] and [12] may serve as application examples.

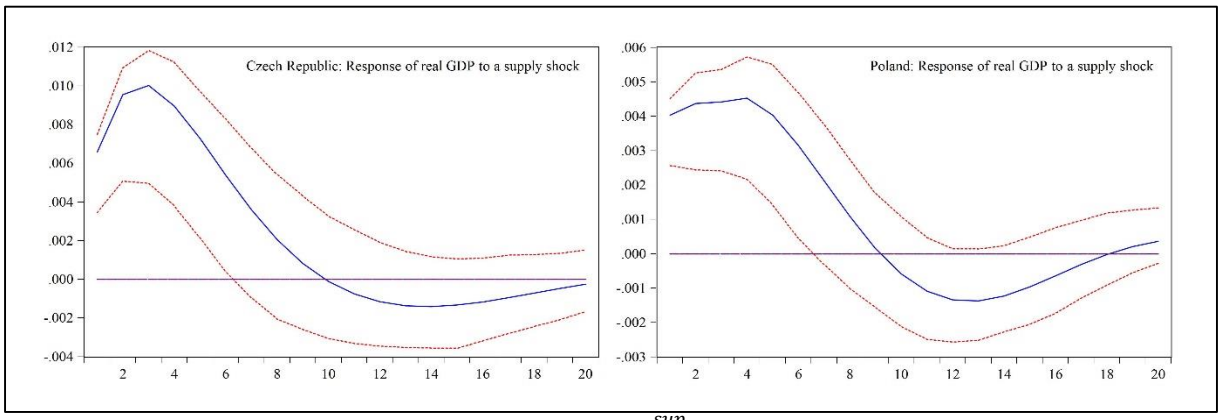


Figure 2 Responses of real GDP to supply shock ε_t^{sup} (with 95% confidence intervals shown)

The estimated reactions of Czech economy to diverse shocks (both positive and negative) may be viewed as key macroeconomic policy factors to be considered during the decision making process. In order to evaluate such reactions on a robust basis, we have estimated the model (4) – (5) using the whole dataset available, as well as after splitting the sample into two sub-samples (2000:Q1 – 2007:Q4 and 2008:Q1 – 2013:Q4) in order to identify significant changes in macroeconomic dynamics resulting from the 2008 crisis. Only mild changes have been identified: Cholesky-identified and BQ-type IRFs generated from the 2008-2013 subsample tend to oscillate more while the effects of shocks fade away. As all data and estimated VARs were stationary, ordinary IRFs converge to zero as time from the original shock increases (thus, cumulative IRFs converge to a new non-zero level, unless zero-value BQ long term identifying conditions apply). Some robustness for such finding was generated through estimating consistently specified VARs and generating IRFs for the same time (sub)samples using "additional" economies: Slovak Republic, Poland and Germany. Although the results obtained through this process provide satisfactory evidence for the robustness (consistency) of our results, the volume of estimation outputs and graphs is overwhelming. Figure 2 provides only a basic illustration of the results, showing the responses of real GDP to a standardized supply shock for Czech Republic and Poland. For both countries, Monte-Carlo generated confidence intervals show significant (non-zero) responses during the first 7 periods only.

Given the accent on robustness, multiple-approach and multiple-model perspective as advocated in [15], we may want to explore the properties of additional and alternatively identified IRFs generated from the estimated VAR equation (4). Most of the Cholesky-based IRF identification scheme results depend on the ordering of variables in a VAR model. Generally speaking, the more contemporaneous correlation exists between residual elements from the estimated VAR equations, the more prominent impact from reordering may be observed on IRFs constructed through Cholesky decomposition. Therefore, low pair-wise correlations between residual series of an estimated VAR model would implicate IRFs that are robust against changes in variables' ordering. We should bear in mind that variables' ordering in a VAR model is usually done on an ad-hoc basis, as VARs are generally defined and constructed as atheoretical. This may easily become a specification-type problem when dealing with IRFs from a high dimensional VAR model with correlated residuals.

In table 1, we provide pair-wise correlations between residuals from the estimated VAR described by equation (4), while observing the ordering of variables in vector $\boldsymbol{y}_t$. Therefore, $\widehat{\boldsymbol{u}}_1$ is a Tx1 $\{t = 1, 2, \dots, T\}$ vector of residuals from the real GDP equation, $\widehat{\boldsymbol{u}}_2$ corresponds to the GDP deflator equation and $\widehat{\boldsymbol{u}}_3$ is the third vector of residuals of the VAR(2) model as described by equation (4). In order to provide additional robustness information, table 1 also includes correlations for residuals from the same model specification, consistently constructed and estimated for Slovak Republic, Poland and Germany. Correlations in table 1 provide mixed information on model (4) robustness against changes in variables' ordering (as far as IRFs from the VAR model are concerned). The authors of this article have experimented with all 6 ordering schemes available for model (4). Using Cholesky decomposition, corresponding IRFs may differ significantly across different specifications (orderings used) of the VAR model (4). We skip detailed results due to space limitations. Data, estimated models and IRFs are available from the authors upon request.

Correlations of residuals	**Czech Republic**	**Slovak Republic**	**Poland**	**Germany**
$corr(\hat{u}_{1t}, \hat{u}_{2t})$	-0.271	-0.155	-0.266	-0.140
$corr(\hat{u}_{1t}, \hat{u}_{3t})$	0.304	0.697	0.203	0.187
$corr(\hat{u}_{2t}, \hat{u}_{3t})$	-0.584	-0.378	-0.650	-0.564

Table 1 Comparison of residuals' correlations from VAR(2) model defined by equation (4)

Also, the authors have made a considerable attempt to estimate the VAR(2) model using panel data methods (specifications with up to four lags were considered with moderate success). Our goal was to transform the increased number of observations available from panel data into improved significance of the estimated model (4) – (5) and better reliability of the generated BQ-type IRFs. In order to avoid the Nickell bias for dynamic panel data models, we use two approaches: First, model (4) – (5) was estimated using data pooling, where all estimated regression coefficients are assumed to be identical across economies (i.e. through pooled estimation, we impose uniform macroeconomic dynamics - which may serve as an approximation to transmission mechanisms - onto the whole panel). Although this approach avoids the Nickell bias altogether, the underlying simplification (restriction) is too strong and the resulting VAR model would not generate statistically significant IRFs. The second and more promising approach is based on a generalized AB estimator. Given its theoretical and practical complexity, this estimation methodology and its applications shall be addressed in a separate working paper.

5 Conclusions

In the paper presented, we address the issue of robustness of the estimated VAR models and IRFs generated using different types of identification schemes. Broadly speaking, relevant and reliable estimates of macroeconomic dynamics and dependencies should be robust - to some reasonable extent - against minor changes to model specification, sample selection and data transformation procedures used. Using quarterly data (2000 - 2013), we evaluate the robustness of real GDP gap (output gap) and inflation gap estimates for use in VAR models. Our finding is that the gap variables calculated (estimated) using HP-filter are relatively robust to the calculation (estimation) method selected and we confirm such results on Czech Republic using macroeconomic data for other countries as well (Austria, Germany, Poland, Slovak Republic). Hence, we may conclude that calculating gap variables (as relative deviations from a steady state) renders good input for a range of econometric models. The resulting series are robust, mostly stationary and bear convenient macroeconomic interpretation.

The input variables described are used for estimation of VAR(2) models and IRFs. Again, we use alternative samples (different time periods as well as economies) and IRF identification schemes in order to evaluate the robustness of our approach and the reliability of the policy selection factors we aim to estimate and evaluate. This paper may serve as a good basis for further macroeconomic evaluation of policy factors, preferably focused on individual policy tools and decision making (influencing) factors.

Acknowledgements

The research project was supported by Grant No. IGA F4/24/2014, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Arellano, M. and Bond, S.: Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, **58**, 1991, 277–297.

[2] Beveridge, S., and Nelson, C. R.: A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'. *Journal of Monetary Economics* **7** (1981), 151–174.

[3] Blanchard,O. and Quah, D.: The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review* **79-4**, 1989, 655–673.

[4] Clark, P. K.: The Cyclical Component of U.S. Economic Activity. *The Quarterly Journal of Economics* **102-4**, 1987, 797–814.

[5] Fernandez-Villaverde, J., Rubio-Ramirez, J. and Sargent, T. J.: *A, B, C's (and D)'s for Understanding VARs*. NBER Technical Working Papers 0308, National Bureau of Economic Research, Inc., 2005, 1–48.

[6] Hušek, R., and Formánek, T.: Srovnání konvergence ekonomik ČR a vybraných zemí eurozóny na základě analýzy funkcí odezvy a nabídkových či poptávkových šoků. *Politická ekonomie* **3**, 2011, 291–309.

[7] Hodrick, R., and Prescott, E. C.: Post War U.S. Business Cycles: an Empirical Investigation. *Journal of Money, Credit and Banking* **29** (1997), 1-16.

[8] Jorda, O.: Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, **95-1**, 2005, 161–182.

[9] Lütkepohl, H.: *New introduction to multiple time series analysis*. New York: Springer, Berlin, 2005.

[10] Nickell, S. J.: Biases in Dynamic Models with Fixed Effects. *Econometrica*, Econometric Society, **49-6**, 1981, 1417–26.

[11] Pesaran, M. and Shin, Y.: Generalized impulse response analysis in linear multivariate models. *Economic Letters* **58-1**, 1998, 17–29.

[12] Shapiro, M. D. and Watson, M. W.: Sources of business cycle fluctuations. In: *NBER Macroeconomics Annual* (Blanchard, O. and Fisher, J., eds.). MIT Press. 1988, 111–148.

[13] Teulings, C. N. and Zubanov, N.: *Is Economic Recovery a Myth? Robust Estimation of Impulse Responses*. CESifo Working Paper Series 3027, CESifo Group Munich, 2010, 1–29.

[14] Wansbeek, T.: On the remarkable success of the Arellano-Bond estimator. *AENORM*, University of Amsterodam, **77-20**, 2012, 15–19.

[15] Wieland, V.: *Model comparison and robustness: A proposal for policy analysis after the financial crisis.* [online] http://www.volkerwieland.com/workingpapers.htm. Working paper, 2011, 1–34.

Using Malmquist index for evaluation of life quality development in the districts of the Czech Republic

Ludvík Friebel[1], Jana Friebelová[1],

Abstract. The main aim of our contribution is an evaluation of life quality in the districts of the Czech Republic during last ten years. As the life quality is depended on many criterions, we had to choose some of multicriteria decision making methods. For evaluation of particular years and districts the Data envelopment analysis (DEA) method was used. This method was initially proposed to evaluate the efficiency. In this paper the level of efficiency represents the level of life quality. In order to assess development of district's efficiency we decided to use Malmquist productivity index (MI). At first we used DEAP SW to compute MI but we had problems with inclusion of uncontrolled inputs into model. Therefore it was necessary to create own application in Maple. The comparison of widely used DEAP with our own procedure was an additional aim of this contribution. The input data set was created based on statistics periodically published by the Czech Statistical Office and Ministry of environment. Our model considered several socioeconomic variables: unemployment rate, criminality, average length of incapacity to work, index of ageing, proportion of economically active inhabitants and specific emissions.

Keywords: data envelopment analysis, Malmquist productivity index, district, LAU1, unemployment, criminality, incapacity of work, index of ageing, dwelling, migration.

JEL Classification: C02, C44, C60
AMS Classification: 90C05, 90B50

1 Introduction

To analyze life quality in particular districts in our work, we have used the Data Envelopment Analysis (DEA) models. This method is advantageous because it does not require initial weights for particular criteria. In this case, the regions were assessed according to the achieved input and output so that the efficiency (the ratio of the outputs and the inputs) would be maximal. Therefore, the potential of the particular regions are respected in the maximal way, see more in section Methodology.

In the paper by authors Martic and Savic [11], the assessment of the regional performance in Serbia was conducted using the DEA Models together with discriminant analysis. To compare effective units between each other, the Andersen-Peterson's Model was used see Andersen and Petersen [1]. The work of Xiong, Liu and Tang [13] shows problems with the choice of criteria for the assessment made with the DEA Method in the field of regional development and the comparison with the static comparative analysis. The social-economic development in the Province of Sichuan is analysed by Li, Cheng [9] using the DEA Method.

We have evaluated the efficiency of 76 districts (LAU 1) in the Czech Republic, according to the selected criteria by the DEA method. This method was initially proposed to evaluate the efficiency. In the first stage, we have considered four types of input: unemployment rate, criminality, average length of incapacity to work, index of ageing and one output: proportion of economically active inhabitants. Because of index of ageing is recognized as uncontrollable, we had to use DEA model with non-discretionary inputs. We evaluated the efficiency of all districts within 11 years (2001-2011). So we can compare their progress with Malmquist productivity index (MI). This index was in its initial formulated by the Swedish economist Sten Malmquist [10] and further developed in the work of Färe and Grosskopf [7].

We tried to involve the economic and social factors into assessment and the reached results were then compared with index of migration effectiveness and price of dwelling.

Because we included 76 districts within 11 time periods in our model, the usage of special DEA software was necessary. The overview of DEA software support is given by Jablonský [8]. At our faculty we have only Banxia Frontier analyst at our disposal. This software can work with basic DEA models including mentioned

[1] University of South Bohemia, Faculty of Economics, Department of Applied mathematics and informatics, Studentská 13, 370 05 České Budějovice, Czech Republic, friebel@ef.jcu.cz, friebelova@ef.jcu.cz.

non-discretionary model, however does not offer calculation of MI. Calculation of MI is possible with DEAP but this software cannot include uncontrolled inputs into model.

2 Methodology

The DEA Models come out from Farrel's Model used to measure the efficiency of the units with one input and one output (Farrel [6]) which was extended by Charnes, Cooper, and Rhodes (CCR) [3] and Banker, Charnes, and Cooper (BCC) [2]. The CCR Models are assumed to have a constant range yield, i.e. the changes of the number of input are proportionally projected to the changes of the number of output. The BCC Models assume variable range yields. The use of the DEA Method is more detailed described, for example, in work of Cooper, Seiford, and Tone [4].

The DEA method is used to divide evaluated subjects (Decision Making Units - DMUs), according to expended inputs and produced outputs, into two groups – efficient and inefficient. The DEA method compares units with the best units on the base of the linear programming theory. In this paper, DMUs are districts of the Czech Republic. Efficiency of the district is conceived as a level of the life quality according to chosen criteria.

Basic DEA models (CCR Charnes, Cooper, and Rhodes [3] and Banker, Charnes, and Cooper (BCC) [2] are either input or output oriented. The output oriented model aims to maximize outputs without requiring a change of one or more of input values. The input oriented model tries to minimize inputs without requiring a change of one or more of output values. In case of inefficient units, the optimal level of outputs or inputs can be determined. The CCR model has assumed that all inputs and outputs can be varied. So we have used modification of this model - CCR model with non-discretionary inputs.

CCR model with non-discretionary inputs

Suppose p DMUs and m inputs (x_i, i =1, 2, …, m), n outputs (y_j, j=1, 2, …, n) for each of these p units. We have to solve p optimizations (one for each of p units) to obtain weight (v) for each of m input and weight (u) for each of n outputs for k-th DMU (k=1, 2,..., p).

Mathematical model for unit H (one of p units) is following linear programming problem:

Maximize

$$\sum_{j=1}^{n} y_{jH} u_{jH} - \sum_{i \in ND} x_{iH} v_{iH} \tag{1}$$

subject to

$$\begin{aligned} \sum_{j=1}^{n} y_{jH} u_{jH} \leq \sum_{i \in ND} x_{ik} v_{iH} + \sum_{i \in ND} x_{ik} v_{iH}\,, k = 1,2, \dots, p, \\ \sum_{i \in D} x_{iH} v_{iH} = 1, \\ u_{jH} \geq \varepsilon, \\ v_{iH} \geq \varepsilon\ (i \in D), \\ v_{iH} \geq \varepsilon\ (i \in ND), \end{aligned} \tag{2}$$

where ($i \in D$) marks inputs, which are discretionary and ($i \in ND$) denotes inputs non-discretionary.

Weights in this model are determined so that objective function (1) is maximal (it is depended on model orientation). If objective function is equal to one, the unit is efficient. An inefficient unit's coefficient is less or more than one (output or input oriented model). For more details see Cooper, Seiford and Tone [4].

The disadvantage of the DEA method, when compared with multicriteria decision making methods, is a certain limitation in terms of the number of inputs and outputs included in the model. It stands to reason that with an increase of inputs and outputs under the same number of assessed unit, the number of efficient unit increases, for more information see Dyson et al. [5]. For this reason, we involved in efficiency assessment only four inputs and one output.

The following data were included into the assessment: unemployment rate (number of unemployed in %), criminality (number of crimes per 10 thousand inhabitants), average length of incapacity to work (number of

calendar days of incapacity to work per one registered event), index of ageing (ratio of inhabitants aged more than 65 to the number of people under 15 - uncontrollable input) and one output proportion of economically active inhabitants (proportion of inhabitants aged from 15 to 64 to the whole population). The source data are available at http://www.czso.cz. By the means of the model with the above mentioned inputs and output, we derived efficiency of particular districts.

In the next stage we computed Malmquist productivity indices based on previously obtained results.

Malmquist productivity index

$$M_o(y_{t+1}, x_{t+1}, y_t, x_t) = \left[\frac{d_o^t(y_{t+1},x_{t+1})}{d_o^t(y_t,x_t)} \times \frac{d_o^{t+1}(y_{t+1},x_{t+1})}{d_o^{t+1}(y_t,x_t)}\right]^{\frac{1}{2}}, \quad (3)$$

where $M_o$is Malmquist productivity index (MI) for unit *o*,
$d_o^t(y_t, x_t)$ and $d_o^{t+1}(y_{t+1}, x_{t+1})$ are distance functions in time *t* and *t*+1, this value corresponds with efficiency of particular unit previously defined in formula (1),
$d_o^{t+1}(y_t, x_t)$ and $d_o^t(y_{t+1}, x_{t+1})$ is previously defined distance functions in time *t*+1 with inputs and outputs in time *t* and $d_o^t(y_{t+1}, x_{t+1})$ is distance funcions in time *t* with inputs and outputs in time *t*+1.

This index is a geometric mean of the period *t* (the first ratio) and period *t*+1 (the second ratio) in order to avoid an arbitrary selection among base years.

We have used own procedure instead of previously mentioned DEAP in order to include one non-discretionary input – index of ageing. We decided for Maple programming environment where simplex algorithm needed for linear programing was already implemented. In order to verify obtained results for randomly selected districts we used Frontier analyst.

In the final stage, we compared obtained MIs within considered period (year 2001-2011) with the average index of migration effectiveness during this period and price of dwelling at the end of this period. We supposed that the efficiency of districts would correspond closely with the mentioned indices. For comparing them we used regression analysis.

Migration effectiveness (*MER*) was first developed by Thomas [12], is defined in the following way:

$$MER_i = 100(D_i - O_i)/(D_i + O_i) \quad (4)$$

Where D_i is total inflows from all other areas, O_i is outflows to all other areas for area *i*.

Migration effectiveness is in other words the ratio between net migration in area *i* and the sum of its total inflows from all other areas and outflows to all other areas.

Price of dwelling was obtained from http://www.realitymorava.cz, where the comparison of dwelling prices in all districts of the Czech Republic is periodically published. With respect to the fact that the average prices of flats are published in different categories, we have derived the average price as a scalar product of the vector representing share the particular category on flats sold and their price vector. The mentioned period covered 13 months – V/2011-VI/2012. The obtained data included 76 districts of the Czech Republic. We had to exclude Prague and from the model, because of its exceptionality originating in the capital status and in the subsidy policy of the EU.

3 Results

In our work, we dealt with life quality on the level of districts (LAU 1). Considering the data accessibility, we assessed regions from year 2001 to year 2011. The source data are available at http://home.ef.jcu.cz/~friebel/research/districts/dea_3.xls. Data were processed in Access.

In Table 1 we can see MIs obtained for ten assessed periods (2001/2002 – 2010/2011) for best ten districts. For evaluating efficiency progress within decade and determining best districts we calculated also geometric mean of MIs for each district. We used geometric mean instead of arithmetical because the MI is geometrical mean between period *t* and *t*+1 too. Seven of ten most efficient districts were the districts with county seat (Brno, Olomouc, Pardubice, Ústí nad Labem, Hradec Králové, Karlovy Vary and Jihlava).

To the contrary Table 2 includes obtained MIs for ten assessed periods (2001/2002 – 2010/2011) for worst ten districts. In this subset we can see especially the border districts. Only two districts (Pelhřimova and Žďár nad Sázavou) don't lie at the border. Geometrical mean values of MI were less than one for all presented regions, see Table 2.

	Year										
District	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**Mean**
Brno-město	0.960	1.001	1.154	0.955	0.813	1.387	1.117	1.189	1.088	1.033	1.060
Olomouc	1.018	0.971	1.079	1.013	1.068	1.006	1.085	1.139	1.078	1.012	1.046
Pardubice	0.952	1.016	1.107	0.991	1.081	1.044	1.079	1.115	1.048	1.027	1.045
Most	1.035	0.983	1.124	0.968	1.037	1.007	1.051	1.227	1.031	0.985	1.042
Ústí nad Labem	0.994	1.000	1.078	0.989	1.006	1.048	1.074	1.170	1.065	0.999	1.041
Hradec Králové	1.012	0.991	1.086	1.016	1.076	1.025	1.123	1.070	0.985	1.022	1.040
Karlovy Vary	0.971	0.993	1.066	0.947	1.083	1.035	1.080	1.121	1.089	1.011	1.038
Jihlava	0.938	0.970	1.091	0.994	1.061	0.985	1.087	1.123	1.120	1.021	1.037
Semily	0.944	1.006	1.084	0.982	1.046	1.018	1.062	1.111	1.092	1.034	1.037
Jičín	0.953	1.030	1.084	0.987	1.087	0.988	1.079	1.105	1.055	0.987	1.034

Table 1 The Malmquist indices for best 10 districts

	Year										
District	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**Mean**
Prachatice	0.980	0.994	0.987	0.995	1.031	0.993	0.981	0.96	0.943	1.019	0.988
Bruntál	0.942	1.024	0.999	0.946	0.981	0.969	0.966	1.085	1.034	0.931	0.987
Žďár nad Sázavou	0.965	0.990	1.003	1.014	0.994	0.964	0.997	0.961	1.020	0.956	0.986
Tachov	0.954	0.982	0.978	1.001	0.969	0.973	0.976	1.042	1.04	0.949	0.986
Vsetín	0.931	1.034	1.023	0.969	0.984	0.96	0.965	1.024	0.975	0.993	0.985
Karviná	0.988	0.968	0.996	0.941	0.960	0.943	1.014	1.103	1.039	0.91	0.985
Pelhřimov	0.908	0.877	1.006	1.126	1.090	1.036	0.931	0.942	0.906	1.058	0.985
Nový Jičín	1.004	0.973	1.017	0.963	0.968	0.972	0.969	0.967	1.021	0.978	0.983
Šumperk	0.967	0.987	0.995	0.937	0.951	0.977	0.967	1.021	0.979	0.954	0.973
Česká Lípa	0.947	0.959	1.004	0.977	0.966	0.977	0.969	0.963	0.987	0.934	0.968

Table 2 The Malmquist indices for worst 10 districts

In Table 3 we can see the comparison of ten best districts from point view of MI and ten best districts form point of view of mean efficiency. This efficiency was computed with CCR model with non-discretionary input (index of ageing) as an arithmetical average within period 2001 – 2011. There we can see that the two sets have no intersection. That finding leads to the assumption that best values of MIs were caused by overall low efficiency in selected districts.

Best MI		**Best Efficiency**	
District	**Mean MI**	**District**	**Mean Efficiency**
Brno-město	1.060	Semily	1.000
Olomouc	1.046	České Budějovice	1.000
Pardubice	1.045	Třebíč	0.997
Most	1.042	Sokolov	0.995
Ústí nad Labem	1.041	Svitavy	0.995
Hradec Králové	1.040	Brno-venkov	0.994
Karlovy Vary	1.038	Rakovník	0.994
Jihlava	1.037	Plzeň-jih	0.981
Semily	1.037	Písek	0.980
Jičín	1.034	Jihlava	0.980

Table 3 Mean Efficiency comparison within 2001-2011 for best districts (MI)

Previously mentioned assumption is indicated in Table 4, where the less efficient districts are presented. Intersection between ten worst districts in terms of efficiency and ten worst districts in terms of MI contains three districts (Karviná, Česká Lípa and Vsetín).

Worst MI		Worst Efficiency	
District	**Mean MI**	**District**	**Mean Efficiency**
Prachatice	0.988	Litoměřice	0.813
Bruntál	0.986	Blansko	0.812
Žďár nad Sázavou	0.986	Jeseník	0.808
Tachov	0.986	Liberec	0.802
Vsetín	0.985	Brno-město	0.798
Karviná	0.985	Kroměříž	0.794
Pelhřimov	0.985	Mělník	0.787
Nový Jičín	0.983	Česká Lípa	0.722
Šumperk	0.973	Vsetín	0.714
Česká Lípa	0.968	Karviná	0.713

Table 4 Mean Efficiency comparison within 2001-2011 for the worst districts (MI)

Hypothesis that migration effectiveness depends on productivity change are tested with OLS regression method. Because of elimination of short-term fluctuation of migration we considered average values in term 2001-2011.This dependency was rejected as we can see in Figure 1. It means that efficiency considered with DEA method with the inclusion of selected inputs and outputs does not affected migration.

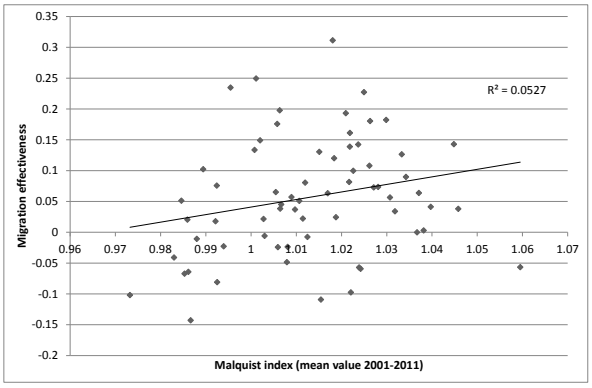


Figure 1 Regression model for migration effectiveness

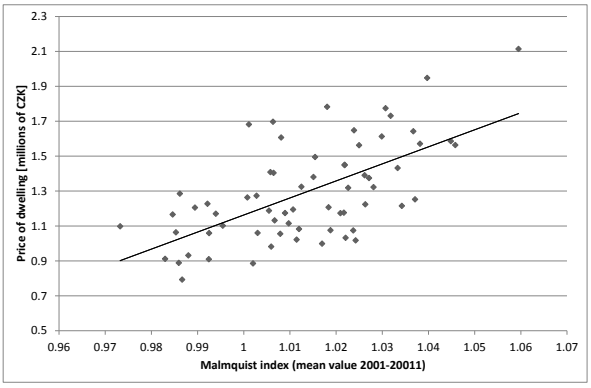


Figure 2 Regression model for price of dwelling

Regarding price of dwelling the situation is rather different. Before applying OLS regression model we had to exclude regions with extreme prices of dwelling from the model. Those districts can be divided into two groups: districts with low price and districts with high price. The firs group contains districts from north bohemian region (Sokolov, Děčín, Chomutov, Louny, Most, Teplice, Ústí nad Labem and Česká Lípa) and one district from north Moravia region – Karviná, in those districts was price of dwelling under CZK 750000. The second group contains districts surrounding the capital - Prague-West and Prague-East, price of dwelling is over CZK 2150000.

As we can see in Figure 2 price of dwelling is depended on MI, correlation index is 0.628. From regression coefficients we can determine following linear function:

$$y' = 9769034x - 8606164,\ I^2 = 0.394 \tag{5}$$

It means that with increasing of mean value of MI by 0.01 price of dwelling increases by CZK 97690.

4 Conclusion

All 76 districts were assessed according to the following criteria: unemployment rate, criminality, average length of incapacity to work, index of ageing (uncontrolable input) and one output proportion of economically active inhabitants. Prague was excluded, see above. We used input oriented CCR non-discretionary model. The same DEA model was used for calculating efficiency in order to determine Malmquist productivity indices.

Two districts are fully efficient within the whole time period – České Budějovice and Semily, but 40 districts were not efficient even once. From point of view of particular years 21 districts were recognized as fully efficient in year 2004.

In terms of efficiency change evaluated with MI among best districts were especially districts with county seat (seven of ten) with unexpended developing potential. Surprisingly among best districts were not districts Prague – West and Prague – East.

Supposed dependency of migration effectiveness on productivity change tested with OLS regression method has not been demonstrated. We have to remember, that people making a decision as where to live have also considered other aspects which are not included in this presented assessment. It means that efficiency considered with DEA method with the inclusion of selected inputs and outputs does not affected migration.

Regarding price of dwelling the situation is rather different. Before applying OLS regression model we had to exclude some regions with extreme prices of dwelling. Index of determination of applied regression was 0.394 and we can see that with increasing of mean value of MI by 0.01 price of dwelling increases by CZK 97690. Thus supposed relation between dwelling price and productivity change was verified.

References

[1] Andersen, P., Petersen, N. C., 1993: A procedure for ranking efficient units in DEA. *Management Science*, Vol. 39: 1261–1264.

[2] Banker, R. D., Charles, A., Cooper, W. W., 1984: Some models for estimating technical and scale inefficiencies in DEA. *Management Science*, Vol. 30, No. 9: 1078–1092.

[3] Charnes, A., Cooper, W. W., Rhodes, E., 1978: Measuring the efficiency of decision making units. *European Journal of Operational Research*, Vol. 2, No. 6: 429–444.

[4] Cooper, W. W., Seiford L. M., Tone K., 2007: *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* New York: Springer, 489 p.

[5] Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., Shale, E. A., 2001: Pitfalls and protocols in DEA. *European Journal of Operational Research*, Vol. 132, No. 2: 245–259.

[6] Farrell, M. J., 1957: The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Vol. 120, No. 3: 253–290.

[7] Färe, R., Grosskopf, S. 1992: Malmquist indexes and Fisher ideal indexes. *The Economic Journal*, Vol. 102: 158-160.

[8] Jablonský, J., 2009: Software Support for Multiple Criteria Decision Making Problems. Management Information Systems, Vol. 4, No. 2: 29-34.

[9] Li, ZG., Cheng, Y., 2010: DEA-Based Evaluation Metod of Harmonious Degree on Socio-Economic Development and Empirical Research. In: *International Conference on Engineering and Business Management – Conference*, Vols. 1–8: 1315–1319.

[10] Malmquist, S, 1953: Index numbers and indifference curves. *Trabajos de Estatistica*, Vol. 4: 209–242.

[11] Martic, M, Savic, G., 2001: An application of DEA for comparative analysis and ranking of regions in Serbia with regards to social-economic development. *European Journal of Operational Research*, Vol. 132, No. 2: 343–356.

[12] Thomas, D.S., 1941: *Social and Economic Aspects of Swedish Population Movements, 1750-1933*. New York: MacMillan.

[13] Xiong, GQ., Liu, YQ., Tang, BQ., 2008: The Evaluation Study of Coordinated Development of Regional Society and Economy Based on DEA/SCA Metod. In: *4th International Conference on Wireless Communications, Networking and MobilenComputing – Conference*, Vols. 1–31: 12748–12751.

Deterministic queuing models in a service line

Martin Gavalec[1], Zuzana Němcová[2]

Abstract. New approach to cost optimization in deterministic linear queuing systems through the modification of their setting is decsribed. The optimization is based on the prediction of the system future development.

Knowledge of the development of the queue lengths during the run of the system is useful for controlling the total costs. A method of finding the optimal setup of service capacities of the servers in each time period in which the service times stay unchanged is suggested in the paper.

Keywords: queueing systems, lengths of the queues, production costs.

JEL classification: C44
AMS classification: 90C15

1 Introduction

The paper describes four types of linear queuing systems depending on the character of requests an service places and suggests the performance optimization method based on the simulation of the system future development. Queuing theory investigates the systems with typical characteristics - there are some units, say servers, which serve some requests coming into the system. These systems have a mass character. We can often meet such systems in a real life - assembly lines, service of the customers in a store, flow of the patients in the hospital, processing the information using the computer, transit of the vehicles through the network etc. [4]

Similar systems were investigated by Heidergott with usage of stochastic methods [1], [2]. The considered queueing systems lead to a special type of max-plus matrices, so called two-diagonal matrices. We follow up this idea with several differences. We consider open linear queuing system and investigate both - the deterministic and stochastic approaches. The usableness of the simplification in case of deterministic approach is examined.

2 Deterministic linear queuing model

Consider an linear queuing system which contains n service places. Incoming requests pass through a series of all servers, then leave the system. Similar system was investigated also in [3].

Every server in the line can serve only one request at a time. The system works in so called stages - it is the time period during which both the service intensity at particular service places and the arrival intensity of the requests to the system are constant. The time variable during stage takes values $t = 1, 2, ..., T$, where T is the length of the stage. There can be arbitrarily long queues in front of each service place before the stage starts.

The termination of the stage can arise of two reasons. The first can be caused by a managerial decision - the change in the setting of the system, and the second is caused from the outside - the change in the arrival intensity to the system.

The intensity of arrivals is represented by the service intensity of the entrance server (indexed by $i = 0$). The server can be considered as a gatekeeper, which controls the access of the requests. It can be assumed, that the queue in front of the entrance server is long enough. Then the gatekeeper admits

[1]University of Hradec Králové/Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 3, 500 03, martin.gavalec@uhk.cz
[2]University of Hradec Králové/Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 3, 500 03, zuzana.nemcova@uhk.cz

the requests to the system with a constant speed - in other words, service time of the entrance server represents the arrival intensity of requests into the system. Let us denote by z_i the basic service time of the server(s) at service place i and K_i as a number of servers at the service place i. For $i = 0$ the change in z_0 (resp. K_0) in the model with the continuous (resp. discrete) service time (explained below) represents external reasons for the change of the stage. Internal reasons are comprised by managerial decisions concerning relevant change in setting at service places $i \in \{1, 2, \ldots, n-1\}$.

Manager of the system can then significantly affect the overall production costs of the system by making the decision in changing service intensity at certain service place.

Production costs P consist of the work costs W, the idle costs I, the costs resulting from long queues Q and the costs for change the service time (or number of servers) on some of service places C, ie

$$P = W + I + Q + C \tag{1}$$

The system manager has the following data available before the system starts: z_i, K_i, queue lengths l_i, the amounts of constituent costs W_i, I_i, Q_i, C_i and maximum tolerated (not yet penalized) queue lengths M_i. In general, there can be K_i identical service providers on each service place, thus the service intensity is $\sigma_i = K_i/z_i$. The values of K_i are natural numbers. The index k denotes particular server at service place i and takes values $k = 1, 2, \ldots, K_i$.

Work costs occur whether the request is being served or not - it is proportionally dependent on the number of the servers K_i. These are the costs necessary to ensure the functioning of servers - for example electricity costs, wage, etc.

Idle costs are costs that are spent when the server is not working. These costs are proportionally dependent on the number of inactive servers at service place i.

Too long queues can cause a loss of profit, and thus also the next sort of costs - queuing costs. This reflects the reality - the longer queue, the bigger dissatisfaction of the customers. The costs resulted from excessively long queues are directly proportional to the degree of exceeding the maximum tolerated queue length M_i.

Change costs are one-shot costs representing amount of money expended for factual change in the system settings. Change costs are added to the the total costs of the stage that have ended by the change.

Possible reaction of manager of the system depends on the ratio of constituent costs.

Let us denote by $l_i(t)$ the length of the queue. The total production costs of the stage are then computed by following formula.

$$P(T,j) = \sum_{t=1}^{T} \left(\sum_{i=1}^{n} W_i K_i + \sum_{\substack{i=1 \\ l_i(t)<0}}^{n} I_i(-l_i) + \sum_{\substack{i=1 \\ l_i(t)>M_i}}^{n} Q_i \left(l_i(t) - M_i\right) \right) + C_j \tag{2}$$

Remark 1. For $K_i = 1$ the value $l_i(t) \geq 0$ denotes the queue of positive length, whereas $-1 < l_i(t) < 0$ denotes that the capacity of the server is not fully used and $l_i(t) = -1$ means, that the server is not working - it is idle. For $K_i > 1$ the value $l_i(t) \geq 0$ also denotes the queue, whereas $l_i(t) < 0$ denotes that the capacity of some of the servers is not fully used or/and some of the servers are idle.

Generally, due to change in the setting of the system, the queues behave differently in each stage. Some of them can remain stable, some of them increase or decrease. In so called transient time, it can happen that the queue is temporarily increasing or decreasing. After the transient time, in which the queues can possibly change the sign of their growth, the stable state is considered. Then the queues do not change their sign of growth i.e. each queue either remains stable, or constantly increases or decreases, except the situation where the queue becomes empty.

We can distinguish four types of queues (according to the behavior in time) - constant queue, decreasing queue, increasing queue and empty queue. The empty queue is a special case of constant queue, or emerges as a result of decreasing queue.

3 Queue length computation

Let us consider that both service times and requests flow, can be of two types - continuous or discrete. If the system with continuous service times is considered, the manager can change the setting of the system by acceleration or deceleration the service times of the servers, so that the condition $z_i > 0$ is fulfilled. Using the discrete service times approach it is possible to add or take away another identical server at some service place in order that $K_i \geq 1$. Continuous requests flow means that there is no need to wait until the whole request is served - the served part of the request can leave the service place and fall in following queue. If the requests flow is discrete, it means that request can fall into next queue soon after its completion. For simplicity we assume that any changes of the service time is performed exactly at one server by one time unit, and all more complex changes are performed as a series of such simple changes. With respect to computational complexity we consider the change costs as constant.

Thus we can distinguish four types of systems: system with continuous service times and continuous requests flow, system with continuous service times and discrete requests flow, system with discrete service times and continuous requests flow, system with discrete service times and discrete requests flow. Each type of the system is described below, together with the basic characteristics needed to compute the total production costs. Characteristics are computed for $i = \{1, 2, \ldots, n-1\}$.

3.1 System with continuous service times and continuous requests flow.

An example of such system is processing the letters at the post office, where requests are represented by packages of letters and every time unit the officer carries away already processed letters from each worker. In case of need workers can slow down or quick up the pace of work.

We suppose that there is one server at particular service place, $K_i = 1$. Then $\sigma_i = 1/z_i$ in this type of system. The given service time z_i can be changed within certain limits. In other words the server can be accelerated or decelerated. We take into account only such deceleration where $z_i > 0$ - the server is still working (it is not possible to put the server out of service).

Because the served part of the request can fall into next queue, we can compute the quantity of already served part of the request.

In further text we use following notation: by $f_i(t)$ we denote the served quantity of request at service place i in time t and by m_i we denote the number of requests present at ith service place, $m_i \geq 0$. Let us also denote by s_i the remaining part of the request that is being served on service place i, $0 \leq s_i \leq 1$. The notation of s_i will be used in the Theorem 2 (for the systems with continuous service time with one server at each service place).

Theorem 1. *Consider a queuing system with continuous service times and continuous requests flow. Then parameters $f_i(t)$, $m_i(t)$ and $l_i(t)$ are computed by formulas:*

$$f_i(t) = \min(m_i(t-1), \sigma_i) \tag{3}$$

$$m_i(t) = m_i(t-1) - f_i(t) + f_{i-1}(t-1) \tag{4}$$

$$l_i(t) = m_i(t) - 1 \tag{5}$$

Proof. The formulas (3), (4) and (5) follow directly from the definition. □

Remark 2. The served quantity is described as a minimum of number of requests in previous time lap and service intensity of ith server. Due to continuous requests flow it can happen, that the capacity of the server is not fully used, $0 < m_i(t-1) < \sigma_i$. It is easily seen, that m_i need not to be whole number.

3.2 System with continuous service times and discrete requests flow.

An example of such system is assembly line manufacturing spare parts for cars. The spare part can move ahead to next service place after its full completion. The speed of machines can be adjusted.

Similarly to the previous type of the system, there is one server at each service place i and the manager can change the service times of the servers. Request flow is discrete - the request remain in the service until it is completed.

Theorem 2. *Consider the system with continuous service times and discrete requests flow. The number of requests present at service place i in time t, $m_i(t)$, and value s_i are then computed by following recursion rules:*

If $(s_{i-1}(t) = 1$ *or* $s_{i-1}(t) = 0)$ *and* $s_{i-1}(t-1) > 0$
  If $(s_i(t) = 1$ *or* $s_i(t) = 0)$ *and* $s_i(t-1) > 0$
    $m_i(t) = m_i(t-1)$; *else* $m_i(t) = m_i(t-1) + 1$; *endif;*
else
  If $(s_i(t) = 1$ *or* $s_i(t) = 0)$ *and* $s_i(t-1) > 0$
    $m_i(t) = m_i(t-1) - 1$; *else* $m_i(t) = m_i(t-1)$; *endif; endif.*
If $m_i(t) > 0$
  If $s_i(t-1) = 0$
    $s_i(t) := 1$; *else* $s_i(t) := s_i(t-1) - \sigma_i$; *endif;*
else
  $s_i(t) := 0$; *endif.* (6)

Proof. Recursion rules follow from the properties of s_i and m_i. □

Theorem 3. *Length of the queue is computed as*

$$l_i(t) = m_i(t) - 1 \tag{7}$$

Proof. The length of the queue is dependent on the number of working servers and there is the only one server at each service place in this type of system. □

3.3 System with discrete service times and continuous requests flow.

Example of such system is similar to example of first described system with continuous service time and continuous request flow - the difference is, that in case of need the manager can add or recall another clerk to some service place.

In case of discrete service times, the manager can change the number of servers on the ith service place, K_i, by adding or taking away some additional server(s) preserving condition $K_i \geq 1$. All servers at service place i have the same service time z_i.

Let us denote by s_{i_k} the remaining part of the request that is being served on service place i on the kth server, $0 \leq s_{i_k} \leq 1$. The notation of s_{i_k} will be used in the Theorems 4 and 5 (for the systems with continuous service time with K_i servers at each service place).

Theorem 4. *Consider the queuing system with discrete service times and continuous requests flow. Then parameters s_{i_k}, $f_i(t)$, $m_i(t)$ and $l_i(t)$ are computed according following formulas.*

If $m_i(t) > 0$
  For $k = 1 : K_i$
    If $s_{i_k}(t-1) > 0$
      $s_{i_k}(t) := s_i(t-1) - \sigma_i$; *endif;*
    If $s_{i_k}(t-1) \leq 0$ *and* $l_i > 0$
      $s_{i_k}(t) := 1$;
      $l_i(t) := l_i(t-1) - 1$; *endif; endfor;*
endif. (8)

$$f_i(t) = \sum_k \min(s_{i_k}(t-1), \sigma_i) \tag{9}$$

$$m_i(t) = m_i(t-1) - f_i(t) + f_{i-1}(t-1) \tag{10}$$

$$l_i(t) = m_i(t) - \sum(s_{i_k}) \tag{11}$$

Proof. The value s_{i_k} is changed according to the service phase. The served quantity of requests, $f_i(t)$, is computed as a minimum of the remaining part of the request and the service intensity, because due to continuous requests flow it can happen, that $s_{i_k}(t-1) < \sigma_i$ (the server is not fully used). Value $m_i(t)$ is computed with respect to incoming and outgoing served parts of requests. The length of the queue is

the difference between the number of present requests and the sum of remaining parts at all servers at the service place. □

3.4 System with discrete service times and discrete requests flow.

Example of such system is similar to the assembly line manufacturing spare parts with the difference in changing the setting of the system. To change the speed of the process at some service place, the another identical machine is added or taken away.

Similarly to the previous type of the system, the manager can change the number of servers at the service place i, however the request flow is discrete.

The remaining part of the request, s_{i_k}, is computed by rules 8.

Theorem 5. *Consider the system with discrete service times and discrete requests flow. Values m_i, l_i are computed by the following rules:*

For $k = 1 : K_{i-1}$
  If $(s_{i-1_k}(t) = 1$ *or* $s_{i-1_k}(t) = 0)$ *and* $s_{i-1_k}(t-1) > 0$
    $x := x + 1$; *endif; endfor;*
For $k = 1 : K_i$
  If $(s_{i_k}(t) = 1$ *or* $s_{i_k}(t) = 0)$ *and* $s_{i_k}(t-1) > 0$
    $y := y + 1$; *endif; endfor;*

$$m_i(t) = m_i(t-1) + x - y. \tag{12}$$

For $k = 1 : K_i$
  If $s_{i_k}(t) > 0$ *or* $s_{i_k}(t-1) > 0$
    $o := o + 1$; *endif; endfor;*

$$l_i(t) = m_i(t) - o. \tag{13}$$

Proof. The value $m_i(t)$ is computed with respect to incoming requests from previous $(i-1)$ service place (variable x) and the completed requests at service place i (variable y). The length of the queue is calculated as a subtraction of working servers (variable o) from the number of present requests at the service place. □

4 Optimization

Optimization is based on finding the appropriate moment to make a change at some service place. The direction of the change (whether the server will be accelerated or decelerated, or whether the server will be added or taken away) is determined by the minimum of average costs of the next (potential) settings of the system.

The function of average costs if the stage ends in time T is denoted by $E(T, j, \delta)^{(r)}$. The index j represents the service place, where the change have been made and the parameter δ denotes the direction of the considered change. In general $\delta \in R$, but for simplicity we suppose the unit changes. Considering the system with continuous service time, the parameter $\delta = 1$ means acceleration of the current service time z_i (analogously the deceleration in case that $\delta = -1$). For the system with the discrete service time the parameter $\delta = 1$ stands for the addition of the server at the jth service place (similarly $\delta = -1$ means taking away the server). Index r stands for the numerical order of the stage.

The function of average costs if the stage ends in time T is computed as follows.

$$E(T, j, \delta)^{(r)} = \frac{P(T, j)}{T} \tag{14}$$

The managerial decision to change the setting of the system is given by the time $t_m^{(r)}$, which is a global minimum of the set of functions $E(T, j, \delta)^{(r)}$ - the functions computed for all possible changes in the rth stage. The type of the change is given by prediction of the behavior of the system in the stage $r+1$. The execution of the change causes that the new stage begins and the decision-making process is repeated.

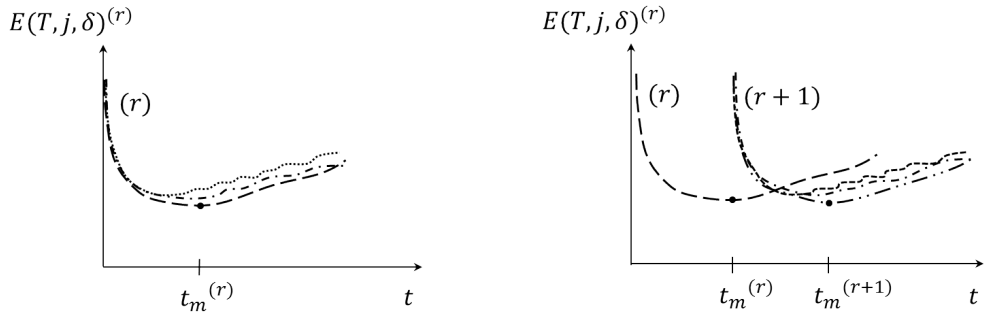


Figure 1 Decision-making process

The decision-making process is illustrated by Figure 1. The curves of the graphs are not smooth. This is a result of the variation in the development of the queues. The functions marked as (r) (left part of the figure) represent the rth stage and were computed by calculating the function of average costs if the stage ends in time T for each possible change. The result of the first step of the decision-making process is the time $t_m^{(r)}$ (global minimum of the set of functions), which answers the question "When to make the change?", because continued work of the system will lead to the growth of the average production costs.

The question "What will be the change like?" is answered by the second step - computation of set of functions marked as $(r+1)$ (depicted in the right part of the figure). These functions simulate the development of average costs for every potential change performed in time $t_m^{(r)}$. The setting of the system will be changed according to the function that has minimum, which is the smallest of all the minims of functions set $(r+1)$. In that moment the next stage begins - parameters of the system were changed. Analogously the time $t_m^{(r+1)}$ determines the most appropriate time to make the next change.

5 Conclusions

Four types of deterministic linear manufacturing systems were described and the method for adapting the setting of the system so that the average production costs are minimal was suggested. The intention of further research is to compare these deterministic models with the stochastic one. For this purpose it is being developed a program for simulation of the system behavior by different settings.

Acknowledgements

The support of the Czech Science Foundation project # 14-02424S is gratefully acknowledged.

References

[1] Heidergott, B.: A characterisation of (max,+)-linear queueing systems. In: *Annual report of the TMR programme ALAPEDES, contract ERB-FMRX-CT96-0074*, European Commission, 1999.

[2] Heidergott, B.: *Max-Plus linear Stochastic Systems and Perturbation Analysis*, The International Series of Discrete Event Systems **15**, Springer–Verlag, Berlin 2007.

[3] Gavalec, M., Němcová, Z. 2011. *Queues in linear manufacturing systems*, Proc. of the 29th Conf. Mathem. Methods in Economics, University of South Bohemia, Jánska Dolina, 176-181.

[4] Jablonský, J.: *Operační výzkum: kvantitativní modely pro ekonomické rozhodování.* Professional Publishing, 2007, 323 p. ISBN 978-80-86946-44-3.

A production line inventory control model with an unbalanced output of the individual production facilities and with respect to other limitations

Lenka Gladavská [1], Lenka Králová [1], Miroslav Plevný [1]

Abstract. Basic models of inventory control deal with a number of simplified assumptions and take no account of the capacity possibilities of a company or the conditions of the supplier-customer relationships. It is necessary to modify these models to be effectively used in practice. Limited storage area or the outsourcing of a part of the manufacturing process as well as specific customer requirements, usually connected with general contracts of delivery, may serve as an example of the limitation to be respected. The aim of the contribution is to show, on a realistic example from practice, a design of a multi-level production line inventory control model that would respect the above conditions with an unbalanced output of the individual production facilities. The model considers namely the in-process production time and the time necessary for creating the reserve of components. Attention is also paid to the calculation of the amount of the material supplies, the adequate amount of the semi-finished and finished products reserves and the safety stock with respect to the limited storage areas.

Keywords: inventory, inventory control model, multi-level production line, limited storage area, internal logistics.

JEL Classification: C44
AMS Classification: 90B05

1 Introduction

Basic models of inventory control deal with a number of simplified assumptions and take no account of the capacity possibilities of a company or the conditions of the supplier-customer relationships. Nowadays, emphasis is put on a system solution of logistic chains considering a number of mutually interlinked aspects that influence the structure and the amount of the total logistic costs and the standard of the provided customer service. It has become necessary to modify the regularly used models not only on the basis of the cost criteria but also with regard to a number of other limitations, such as specific requirements of customers usually defined by the general delivery contracts, outsourcing, limited storage areas, the quantity and capacity of the means of transport, the quantity and capacity of transportation and manipulation units or the minimum size of batch in relationship to the production facility given by the technological requirements for the manufacturing of a given product.

The aim of the contribution is to show, on a realistic example from practice [10], a design of a multi-level production line inventory control model that would respect the above conditions with an unbalanced output of the individual production facility. The model considers namely the in-process production time and the time necessary for creating the reserve of components. Attention is also paid to the calculation of the amount of the material supplies, the adequate amount of the semi-finished and finished products reserves and the safety stock with respect to the given limitation. The described model ensures harmony of the internal logistics and the manufacturing process with the requirements of such company customers whose orders are not equally placed.

2 Literature review

Stochastic models dealing with the demand and delivery terms of random character are used in practice. To deal with insignificant fluctuation of demand for a certain item or in case of more complicated models of stochastic character it is possible to use deterministic models [5]. EOQ (Economic Order Quantity) and EPQ (Economic Production Quantity) models and their modifications are most often used for planning material requests with the aim of setting the batch sizes on the production facility of production lines. The original EOQ [9] and EPQ [12] models were later extended by the losses from the prematurely exhaustion of supplies. Lately, algebraic optimization methods instead of the models on the basis of differential equations have been preferred in the systems of inventory control, the reason being their solvability. Grubbstorm and Erdem [8] used an algebraic procedure for the EOQ model with backorders. Cárdenas-Barrón [2] suggested algebraic optimization of a

[1] University of West Bohemia, Faculty of Economics, Husova 11, Plzeň, Czech Republic, gladavsk@kem.zcu.cz

supply shortage EPQ model and in [3] an EOQ/EPQ model with two kinds of backorder costs. Ronald et al. [11] simplified the procedure of Grubbstorm and Erdem [8] and Cárdenas-Barrón [2]. Most of the classic production inventory systems deal with the economic order quantity of materials used for the manufacturing of a product and the economic production quantity of a given product separately. There are, however, also models that combine the optimization of materials and of a product. Goyal [7], for example, or Balkhi [1] present integrated production inventory systems – Goyal [7] for single product, Balkhi [1] for more final products. These models do not, however, take into account various limitations that may arise in practice when controlling the supplementary supplies on production lines, or they only consider very simple limitations. García-Laguna [6], for example, uses the adapted EOQ and EPQ models where the extent and number of orders must be an integer. In the following text we are going to illustrate some possibilities of modeling the supplementary and safety stock in a production line when, at the same time, respecting some limitations that often occur in practice.

3 A model of inventory control in a production line

In the following text we are going to deal with a model of a multi-level production line in which there are both the facilities operating in the batch mode and the facilities operating continuously. In our considerations we are going to limit ourselves to the production of only one product. Our aim was to formulate a simple procedure of the cost optimization of inventory control on a production line with respect to the given limitations. In the inventory control model it is necessary, in particular practical cases, to respect various limitations, such as for example the following ones:

- conditions based on the supplier-customer relationships;
- limited storage areas between the individual production facilities;
- quantity and capacity of the transportation and manipulation unit;
- quantity and capacity of the means of transport at the input and output of the production line;
- outsourcing of a part of the production process;
- technical and technological parameters of the production facilities influencing the minimum and/or the maximum size of the production batch.

The conditions of the supplier-customer relationships *are usually stipulated* in the form of the so called general contracts, where all the requirements for future purchases are usually defined for a given time period. Within such a period there is usually no guarantee of equal distribution of purchases. Another possibility in concluding general contracts is to define an assumed amount of purchases which can be modified later in the fixed dates. The closer the delivery date is the smaller the permissible deviation from the originally set value.

The limited storage areas set various upper limits of the supplementary inventory. Similarly, the quantity and capacity of the relevant transportation and the manipulation units can limit the amount of inventory between the individual production facilities. On the contrary, the limitations for the size of a batch in a production facility are set by the lower limits, i.e. the minimum quantity that is possible and also effective to be produced in one batch. Further in the text we are going to analyze the principles of drawing up models for inventory control on a production line in more detail. We are going to consider a simple model of a production line with unbalanced performance of a production facility which starts by the input of material from an external supplier to the stock of material. The production takes place in two stages; the first facility operating in the batch mode passes the semi-finished product to the next workplace which operates continuously (see Fig. 1).

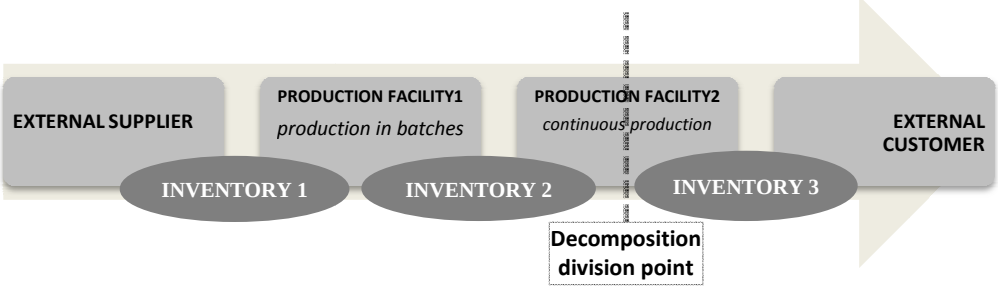


Fig. 1 Schematic illustration of a simple production line

In case of modelling more complicated real systems it is advantageous when the system may be **decomposed** into several smaller units which can be dealt with autonomously. In case of modelling the inventory on a production line with more production facilities it is possible to find suitable **division points** for the decomposition **in the place of the continuously operating production facilities**. PRODUCTION FACILITY 2 in Fig. 1 can serve as an example of such a point. Thus the only standard condition related to this point is ensuring enough inputs for this continuously operating facility. On the side of the output the products with constant intensity are generated regardless of anything that was happening in the production line before this point. Modelling of the extent of INVENTORY 3 need not directly consider the way of the solution of INVENTORY 2.

The type (stochastic or deterministic) of the applied model of inventory control is usually determined by the conditions of the supplier – customer relationships. According to Fiala [5] it is, in case of an insignificant fluctuation demand for a certain item or in case of more complicated models of stochastic character, possible to use the deterministic models as an approximation solution. The above mentioned **general contracts** are another relevant reason for **applying simpler deterministic models** in the real situations. In these contracts the duty of the supplier to always meet the requirement of the customer is already embedded, i.e. it is possible to deduce **the worst possible scenario**. Provided we specifically know the worst possible case from the point of view of the deliveries to the customer, the following pattern can be used for modelling the progress of an inventory:

- ***the deterministic model*** respecting the standard progress of the production development combined with
- ***the safety stock*** covering the worst possible case.

For the deterministic modelling of the progress of inventory without admitting the lack of inventory the models EOQ (Economic Order Quantity) and EPQ (Economic Production Quantity) are used most often. The EOQ model is often used for modelling the inventory at the input into the production line when the input material is fed in batches (see INVENTORY 1 in Fig. 1). The EPQ model can be used in the production line for modelling the supplementary stock immediately after the facility operating in the batch mode (see INVENTORY 2 in Fig. 1). These classic models do not count on further limitations we mentioned above.

The progress of the stock level at the "end" of the production line may be a specific case of inventory control on production lines where it may happen that the inventory is evenly replenished from the last link of the production line and, at a certain moment, it is dispatched to the customer at once (as a batch) (see INVENTORY 3 in Fig. 1). This progress of the stock level is analogous with the EOQ model with the only difference that the inventory grows evenly and it falls at once instead of being replenished at once and after that evenly consumed (see Fig. 2). For this reason we will call this model further in the text the **"reversed" EOQ model**.

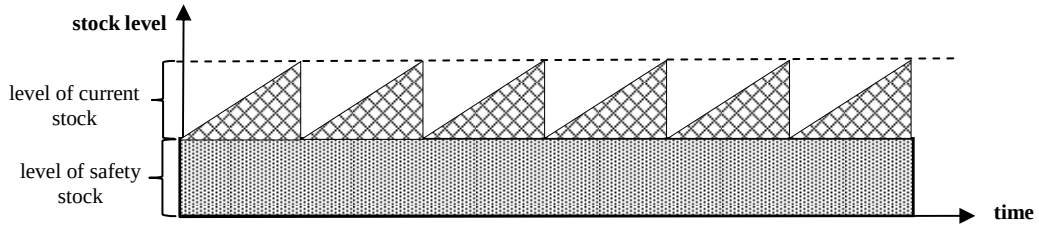


Fig. 2 The "reversed" EOQ model with considering the safety stock

In considering the way of calculation of the safety stock it is generally necessary to consider the variability of the demand and of the delivery terms:

1. the amount of the **material stock** (INVENTORY 1) is influenced mainly by the delivery terms;
2. the level of the **stock of the semi-finished product** (INVENTORY 2) after the production facility operating in the batch mode must respect the fact that in this facility also other products are manufactured; therefore it is not possible to interrupt its production immediately and to satisfy only the demand of the analysed product;
3. the safety **stock of the finished product** (INVENTORY 3) must be maintained because of the high probability of irregular calls from the side of the customer during one period of the monitored time.

Generally, the amount of the safety stock depends on the required reliability of the deviation coverage. At the same time we must respect the very strict conditions from the side of the customer which do not admit any drop out in the deliveries of the final product. For the calculation of the safety stock we therefore recommend a simple pattern in the form of the product of the required quantity and the time necessary to remove the obstacles of the standard progress of the manufacturing process.

Ad 1) Material stock

In practice there are two basic ways of dealing with ensuring the necessary quantity of material at the input. In case of implementing the JIT deliveries the volume of INVENTORY 1 in Fig. 1 is not necessary to be considered. In current cases the period of deliveries of material is significantly longer than the production period of the input production facility operating in the batch mode. In such cases the optimum amount of the order of this material from an external supplier may be solved by means of the well-known Harris-Wilson (Camp) formula for the EOQ model [9]. At the same time it is necessary to consider the space capacities for storing the material, or, as the case may be, the capacity of the transportation means. The space determined for storing the material must accommodate the maximum quantity of the material stock, i.e. ideally the optimum quantity of the order and, at the same time, the volume of the safety stock Z_p^m kept for the case of fluctuations in delivery terms (1).

$$Z_p^m = Q_d^m \, t_{opr}^{dod} \tag{1}$$

In relation (1) Q_d^m is the average consumption of the material after the time unit (one day) and t_{opr}^{dod} is the time necessary for the production renewal in case of the supplier (in days).

Ad 2) The stock of the semi-finished product

For the above mentioned reasons a production-consumption model of the periodically replenished stock, well-known under the abbreviation of EPQ, may be used for INVENTORY 2. The stock of the semi-finished product resulting from the previous production facility with production rate K^p is stored periodically in the storage area from which the follow up production facility draws. This stock is principally influenced by the **size of the production batch in the first production facility** d^{opt} which is given mainly by the time for preparation and finishing of production and the average demand rate Q^p of the follow up production level for a given time period. The way of calculation of the optimum size of the production batch d^{opt} is generally very well-known and it can be found for example in [4, p.117] as $d^{opt} = \sqrt{\left[2Q^p c_b / (c_s T)\right] \cdot \left[K^p / (K^p - Q^p)\right]}$, where c_b and c_s are the batch and storage costs.

When considering the **lower limits for the size of the production batch** d_{min} (2) it is necessary to respect partly the potential requirement for the minimum size of the production batch based on using the production technology, and partly the frequent requirement of the maximum permissible ratio of the time of idleness (the time for the preparation and finishing of the production operations t^p_{pz}) and the overall disposition period of this facility. The period of the active operation can be expressed by a multiple of the size of one production batch d^p_v and time t^p_k for carrying out the production operation on one unit of the processed product.

$$d_{min} = \frac{t^p_{pz}\left(1 - r^{max}\right)}{r^{max}\, t^p_k} \tag{2}$$

The size of the production batch is often limited by the **maximum storage capacities** for the semi-finished products (i.e. the maximum stock and the safety stock). The maximum stock of the semi-finished products x^p_{max} is influenced by the production capacity of the first production facility and the consumption of the follow up production facility, and by the size of the production batch t^p_d, see the relation (3) adapted by [4, p.117].

$$x^p_{max} = t^p_d \frac{K^p - Q^p}{T} = \left(1 - \frac{Q^p}{K^p}\right) d^{opt} \tag{3}$$

The safety stock of the semi-finished product Z^p_p must be kept because of the fluctuation of the demand for other products processed in the first facility, or possibly because of its potential malfunctions in the future. In the relation (4) the value t^1_{opr} represents the time in which a company is able to deal with a problem in the first production facility, for example with technical malfunctions or the necessity to manufacture other products.

$$Z^p_p = Q^p \times t^1_{opr} \tag{4}$$

Ad 3) The stock of the finished product

In looking for the adequate **size of the finished product** we have to respect the provisions based on the general contract with the customer. The size of his/her order is usually limited by the maximum value for a certain period (for example a week or a month). If we consider the worst possible variant, we have to keep a safety stock of the finished products at the level that covers the maximum height of the call of the customer. The overall extent of the stock of the finished product will be formed by the safety stock in the extent of the maximum allowed order from the side of the customer and the current stock continuously generated by the production line.

In the previous text we did not deal with the limits given by the storage areas at all levels of the production process. With regard to the fact that these areas are usually firmly given, it is necessary to compare the calculated optimum quantity of the stock with the upper and possibly even with the lower limits. If the optimum size of stock at the individual production level did not correspond to these limits, it would be necessary to modify the size of stock accordingly. With regard to the fact that the considered cost functions are usually convex, this modification consists in respecting the closest possible permissible value of the given interval and in the adequate modification of the size of the production batch. It is not, however, possible to reduce the size of the safety stock of the finished stock without the modifications of the general contract.

4 A practical example

Let us consider the production of a component for car industry which is manufactured from clips that are imported from an external supplier and before the production itself they are stored in the store of material. The clips are worked in batches on a stamping press and then they wait in the form of the semi-finished product for the process of clinching in the second production facility operating in the non-stop mode. From there they are transported to the store of the finished product where they wait for the call from a customer (see Fig. 3). With regard to the continuous character of the production in the second production facility (clinching machine) this facility can be considered to be the **division point of the decomposition** of the production line (see Section 3). The amount of the inventory in the production line can be solved in two independent steps: 1) the optimization of the stock of material and the stock of the semi-finished product; 2) setting the stock of the finished product.

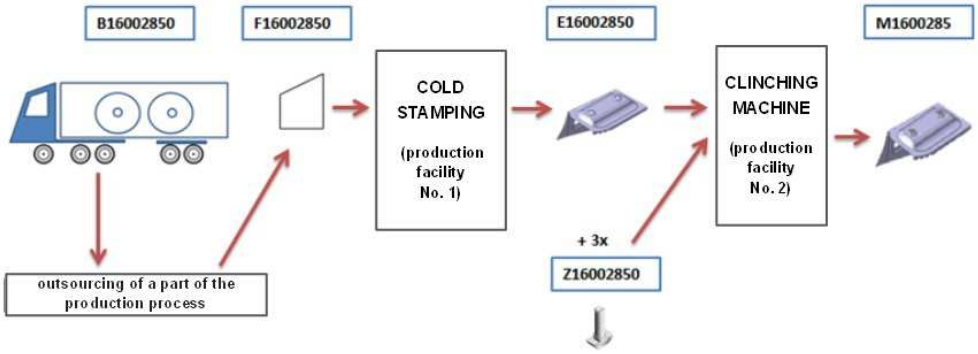


Fig. 3 The scheme of the production process in the production line

In the model it is necessary to respect the following limitations:

a) ***Conditions of the supplier-customer relationships*** – the general contract with the customer is set to the year´s call of 550,000 pieces, while the average weekly call amounts to 11,000 pieces. The customer sends the call one working day ahead. Within the weekly call the required goods can change by up to 60% from the average weekly call - i.e. 11,000 ± 6,600.
b) ***Outsourcing of a part of the production process*** – the clips are delivered on half-pallets (0.8 x 0.6m) with capacity 250 pcs, the weight of one piece is 1.141 kg.
c) ***Limited storage areas for the input material*** 37.5 m^2 enables realistic storing of maximum 72 half-pallets.
d) ***Limited storage areas for the semi-finished product*** 25 m^2 enables realistic storing of maximum 60 Euro-boxes (1.240 x 0.835 m) that are stackable into three layers and can accommodate 100 pcs of semi-products.
e) ***Limited storage areas for the finished product*** 49.5 m^2 enables storing of maximum 264 customer boxes (0.8 x 1.2 m) with capacity of 50 pieces of products that are stackable up to six layers.
f) ***Quantity and capacity of the transportation and manipulation units*** – the manufacturer together with the customer have an agreed fixed stock of boxes maintained by the manufacturer in the extent of 220 pcs.
g) ***Quantity and capacity of the means of transport*** – the external supplier trucks with the capacity of 26 tonnes are used for the transportation of clips; the delivery terms Incoterms EXW are agreed with the customer and therefore the external transport at the end of the production process does not represent any limiting factor.
h) ***Technical and technological parameters of the production facility*** – the capacity of the first production facility: 3,000 pcs per shift; the exchange of forms in the press: 40 min on average (the company insists that this time must not exceed 10% of the disposition time of the facility); the clips on the press are put into the facility directly on the manipulation units (half-pallets with 250 pcs) therefore the production batch must be a integer multiple of 250 pcs; the production capacity of the clinching machine is 900 pcs/shift (2,700 pcs/day).

The costs for one shipment in the amount of CZK 12,500 per truck, and the annual costs for maintaining one item of the stock material in the amount of 30% of the value of the stored item (CZK 45 per piece) were calculated by the company. One year is considered to be the monitored period, i.e. 365 days. It was found empirically that in case of a breakdown of the production facilities the supplier is able start the production again within two days at the latest. In case of the standard 5 day working week and the average call of 11,000 pcs it is possible to determine ***the safety stock*** according to (1) as Z^m_p = 4 400 pcs. For storing the safety stock of the material it is necessary to use 18 half-pallets. All the results are summarized in Tab. 1.

Cost-optimum amount of order	**Limitations**	**Order adjusted to limitations**
31,914 pcs i.e. 36,414 kg	Capacity of vehicle	22,787 pcs
	Entire storage area	18,000 pcs (=72*250)
	Storage area without safety stock	**13,500 pcs** (54 halfpallets)

Tab. 1 The cost-optimum amount of the order of the clips respecting the given limitations

Cost-optimum size of one production batch **by EPQ**		**5,457 pcs**
Limitations:	• multiples of 250 pcs (see h))	*5,500 pcs*
	• lower limit for the production batch d_{min} according to (2)	*2,250 pcs*
	• maximum stock x^p_{max} according to (3)	*4 125 pcs*
	• the maximum storage area capacity (see d))	*60 Euroboxes à 100 pcs*
	• the safety stock (one day production of the facility No. 2)	*2,700 pcs (27 Euroboxes)*
	• the storage area capacity without safety stock	*33 Euroboxes à 100 pcs*
Max amount of the production batch (according to (3) where x^p_{max} = 3,300 pcs)		**4,400 pcs**
Limitations:	• multiples of 250 pcs (see h))	*4,250 pcs*
Cost-optimum size of the production batch with regard to the given limitations		**4,250 pcs**

Tab. 2 The cost-optimum size of the production batch on the press respecting the given limitations

When calculating **the cost-optimum size of the production batch on the press** it is necessary to respect the limitations a) to h), and also the costs for preparation and finishing of the production (CZK 9,302) and the costs for storing one item (CZK 8.93) - both calculated by the company. The results are summarized in Tab. 2.

Within **the second step** (after the division point of the decomposition of the production line) it is necessary to determine **the cost-optimum safety stock of the finished product** (see Tab. 3).

Requirement of the customer (see a))	11,000 pcs
Safety stock of the finished product (requirement minus 1 day production – see a))	**8,300 pcs**

Tab. 3 Calculation of the safety stock of the finished product

As is obvious from the above calculations the figures shown in Tab. 4 will be the **cost-optimum parameters for the company with regard to all the limitation conditions**.

Size of	**Stock of clips (pcs)**	**Cold stamping (pcs)**	**Clinching machine (pcs)**
Order/production batch	**13,500**	**4,250**	operates continuously
Safety stock	**4,400**	**2,700**	**8,300**

Tab. 4 The resulting cost-optimum values for the company with regard to all the limitations

5 Conclusion

In the paper we presented a model of inventory control in a production line with unbalanced performance of production facilities. This model respects limitations that often occur in practice including the conditions based on the supplier-customer relationships, the limited storage area between the individual production facilities, the quantity and capacity of the transportation and manipulation units, the quantity and capacity of the means of transport at the input or output of a production line, the outsourcing of a part of the production process and the technical and technological parameters of the production facilities influencing the minimum size of the production batch. In a model demonstrated on a realistic example from practice a possibility of the **decomposition of a model of inventory control in a production line** into smaller and simpler partial models is shown. Setting the cost-optimum amount of the material supply, an adequate extent of the semi-finished and finished product and that of the safety stock is shown when respecting a number of the above mentioned limitations. The presented model considers the current production time and the time necessary for creating the necessary stock of components and ensures the harmony of the internal logistics and the production process with the requirements of the customers whose orders are not evenly distributed.

References

[1] Balkhi, Z. T: On the optimality of multi item integrated production inventory systems. In: *Proceedings of the 3rd Int Conf on Applied Mathematics, Circuits, Systems, and Signals*. 2009, 143-154.

[2] Cárdenas-Barrón, L.E: The economic production quantity (EPQ) with shortage derived algebraically. *Int. J. Prod. Econ.* **70** (2001), 289–292.

[3] Cárdenas-Barrón, L.E: The derivation of EOQ/EPQ inventory models with two backorders costs using analytic geometry and algebra. *Applied Mathematical Modelling* **35** (2011), 2394-2407.

[4] Daněk, J. and Plevný M.: *Výrobní a logistické systémy*. Západočeská univerzita v Plzni, Plzeň, 2005.

[5] Fiala, P.: *Modelování a analýza produkčních systémů*. Professional Publishing, Praha, 2002.

[6] García-Laguna, J., San-José, L. A., Cárdenas-Barrón, L. E., Sicilia, J.: The integrality of the lot size in the basic EOQ and EPQ models: Applications to other production-inventory models. *Applied Mathematics and Computation* **216** (2010), 1660-1672.

[7] Goyal, S.K.: An integrated inventory model for single product system, *Operations Research Quarterly* **28** (1977), 539-545.

[8] Grubbstrom, R.W. and Erdem, A.: The EOQ with backlogging derived without derivatives. *Int. J. Prod. Econ.* **59** (1999), 529–530.

[9] Harris, F.W. : How many parts to make at once factory. *Mag. Manage* **10** (1913), 135–136 and 152.

[10] Kopalová, Š.: *Material Flow management in a particular company*. Bachelor Thesis, Faculty of Economics, University of West Bohemia in Pilsen (2013).

[11] Ronald, R.J., Yang, G.K., Chu, P.: Technical note – the EOQ and EPQ models with shortages derived without derivatives. *Int. J. Prod. Econ.* **92** (2004), 197–200.

[12] Taft, E.W.: The most economical production lot. *Iron Age* **101** (1918), 1410–1412.

[13] Tomek, G. and Vávrová, V.: *Řízení výroby a nákupu.* Grada Publishing, Praha, 2007.

On the consistency of an estimator for hierarchical Archimedean copulas

Jan Górecki[1], Marius Hofert[2], Martin Holeňa[3]

Abstract. The paper addresses an estimation procedure for hierarchical Archimedean copulas, which has been proposed in the literature. It is shown here that this estimation is not consistent in general. Furthermore, a correction is proposed, which leads to a consistent estimator.

Keywords: Hierarchical Archimedean copula, Kendall distribution function, parameter estimation, structure determination, consistency

JEL classification: C51, C46
AMS classification: 62H99

1 Introduction

Hierarchical Archimedean copulas (HACs), which generalize Archimedean copulas (ACs) [13, p. 109], constitute a popular class of copulas, which has been used in high-dimensional applications. There exist successful applications of HACs in finance, e.g., for pricing collateralized debt obligations; see [4, 8]. Given some data, a HAC model can be build using several estimation techniques. Concerning those techniques, one can see that the estimation of a HAC can generally be divided in two subtasks: 1) structure determination and 2) parameter estimation. Both tasks are connected and hence, if the structured determination is performed poorly, i.e., if the true structure is not determined, then parameter estimation also leads to a poor result.

We are aware of two papers [2, 14] addressing both subtasks. The paper [14] describes a multi-stage HAC estimation procedure, in which the HAC structure is determined iteratively in a bottom-up manner. The estimation of the parameters is mainly performed using the maximum-likelihood (ML) technique, but the authors also briefly mention an alternative, which uses the relationship between the copula parameter and the value of Kendall's tau computed on a bivariate margin of the copula (shortly, $\theta - \tau$ relationship). The experimental results on simulated data show acceptable performance for the low-dimensional (5 variables) copula models considered in the paper. However, for some specific copula models, the quality of the results given by the mentioned procedure may be poor, as is addressed and experimentally shown in [2]. The paper [2], which empirically extends the theoretical procedure proposed in [1], also describes an alternative estimation procedure based on the $\theta - \tau$ relationship that can be successful in general.

Based on the research presented in the above-mentioned papers, we aim to show that, if the procedure presented in [14] is properly adjusted, the resulting estimator turns from a generally inconsistent estimator to a consistent estimator of the copula parameters.

The paper is structured as follows. Section 2 recalls some necessary theoretical concepts concerning copulas in general, ACs and HACs. Section 3 addresses the problem of the HAC estimation procedure described in [14]. Section 4 presents our proposed adjustment of the procedure and Section 5 concludes.

[1]Department of Informatics, SBA in Karvina, Silesian University in Opava Karvina, Czech Republic, gorecki@opf.slu.cz
[2]Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada, marius.hofert@uwaterloo.ca
[3]Institute of Computer Science, Academy of Sciences of the Czech Republic, Praha, Czech Republic, martin@cs.cas.cz

2 Preliminaries

2.1 Copulas

Definition 1. For every $d \geq 2$, a *d-dimensional copula* (shortly, *d-copula*) is a d-variate distribution function on $\mathbb{I}^d$, $\mathbb{I} = [0,1]$, whose univariate margins are uniformly distributed on $\mathbb{I}$.

Copulas establish a connection between a distribution function (df) H and its margins $F_1, \ldots, F_d$ (we use the term *margin* as an equivalent for *univariate margin*) via $H(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d))$, $\mathbf{x} \in \mathbb{R}^d$, as is well-known by Sklar's Theorem; see [15]. $H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$, where $F_1, \ldots, F_d$ are the margins of H. In case the margins $F_1, \ldots, F_d$ are all continuous, C is uniquely given by $C(u_1, \ldots, u_d) = H(F_1^-(u_1), \ldots, F_d^-(u_d))$, where $F_j^-, j \in \{1, \ldots, d\}$, denotes the pseudo-inverse of F_j given by $F_j^-(u) = \inf\{x \in \mathbb{R} \mid F_j(x) \geq u\}$, $u \in \mathbb{I}$. As an example, elliptical copulas are derived in this way from multivariate elliptical dfs.

2.2 Archimedean Copulas

Archimedean copulas (ACs) are not constructed using Sklar's Theorem, but instead, one starts with a given functional form and asks for properties needed to obtain a proper copula. As a result of such a construction, ACs are expressed in closed form, which is one of the main advantages of this class of copulas. For construction of ACs, we need the following notions; see [12] for a more general construction of ACs.

Definition 2. An *Archimedean generator* (shortly, *generator*) is a continuous, nonincreasing function $\psi\colon [0,\infty] \to [0,1]$, which satisfies $\psi(0) = 1$, $\psi(\infty) = \lim_{t\to\infty} \psi(t) = 0$ and is strictly decreasing on $[0, \inf\{t \in [0,\infty) \mid \psi(t) = 0\}]$. We denote the set of all such functions by Ψ.

Definition 3. A function f is called *completely monotone* (shortly, c.m.), if $(-1)^k f^{(k)}(x) \geq 0$ holds for every $k \in \mathbb{N}_0$, $x \in [0,\infty)$. We denote the set of all completely monotone generators by Ψ_∞.

Definition 4. Any d-copula C is called *Archimedean copula*, if it admits the form

$$C(\mathbf{u}) = C(\mathbf{u}; \psi) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)), \ \mathbf{u} \in \mathbb{I}^d, \tag{1}$$

where $\psi \in \Psi$ and the $\psi^{-1}\colon [0,1] \to [0,\infty]$ is defined $\psi^{-1}(s) = \inf\{t \in [0,\infty) \mid \psi(t) = s\}$, $s \in \mathbb{I}$.

A condition sufficient for C to be a copula is stated as follows.

Theorem 1. [10] If $\psi \in \Psi_\infty$, C given by (1) is a copula in any dimension d.

2.3 Hierarchical Archimedean copulas

A copula is called *hierarchical Archimedean* if it is an AC with arguments possibly replaced by other hierarchical Archimedean copulas. In this paper, we consider the working example

$$C(\boldsymbol{u}) = C_0\big(u_1, C_1(u_2, u_3), C_2(u_4, u_5, C_3(u_6, u_7, u_8))\big), \tag{2}$$

where the *sector copula* C_k denotes an AC with generator ψ_k, $k \in \{0,1,2,3\}$. Let us mention at this point that all presented ideas can be extended to arbitrary HACs (possibly at the cost of a significantly more complicated notation). Figure 1 shows the corresponding tree representation of this HAC.

According to *nesting condition* presented by [9, p. 88] and [11], nesting ACs leads to a proper copula if all nodes of the form $\psi_k^{-1} \circ \psi_l$ appearing in the hierarchical structure have completely monotone derivatives. In what follows, we assume this sufficient condition to hold. For a list of generators that fulfill this condition, see, e.g., [4, p. 65, pp. 115] or [5].

A random vector $\boldsymbol{U}$ distributed according to an HAC allows for a simple stochastic representation; see [6]. A random vector $\boldsymbol{U}$ distributed according to the HAC (2), e.g., can be represented as

$$\boldsymbol{U} = \left(\psi_0\Big(\frac{E_1}{V_0}\Big), \psi_1\Big(\frac{E_2}{V_{01}}\Big), \psi_1\Big(\frac{E_3}{V_{01}}\Big), \psi_2\Big(\frac{E_4}{V_{02}}\Big), \psi_2\Big(\frac{E_5}{V_{02}}\Big), \psi_3\Big(\frac{E_6}{V_{23}}\Big), \psi_3\Big(\frac{E_7}{V_{23}}\Big), \psi_3\Big(\frac{E_8}{V_{23}}\Big)\right), \tag{3}$$

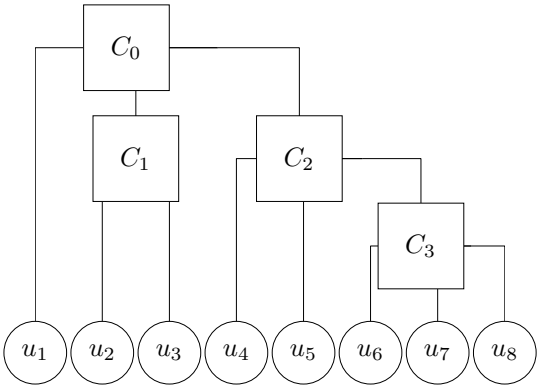


Figure 1 Tree structure of the HAC C as given in (2).

where E_j, $j \in \{1, \ldots, 8\}$, are i.i.d. random variables with standard exponential distribution Exp(1), independent of the random variables V_0, V_{01}, V_{02} and V_{23}. The most important ingredients of the stochastic representation (3) are the random variables V_0, V_{01}, V_{02} and V_{23}. The dependence among these random variables determines the tree structure of the HAC. It can be described as follows. First, draw $V_0 \sim F_0(x) = \mathcal{LS}^{-1}[\psi_0](x)$, where $\mathcal{LS}$ denotes Laplace-Stieltjes transform, i.e., V_0 is distributed according to the distribution whose Laplace-Stieltjes transform is ψ_0. Next, draw $V_{01}|V_0 \sim F_{01}(x; V_0) = \mathcal{LS}^{-1}[\psi_{01}(\cdot\,; V_0)](x)$, where $\psi_{01}(t; V_0) = \exp\big(-V_0\psi_0^{-1}(\psi_1(t))\big)$. Similarly, draw $V_{02}|V_0 \sim F_{02}(x; V_0) = \mathcal{LS}^{-1}[\psi_{02}(\cdot\,; V_0)](x)$, where $\psi_{02}(t; V_0) = \exp\big(-V_0\psi_0^{-1}(\psi_2(t))\big)$. Finally, on the innermost nesting level of the copula C in (2), the random variable V_{23} resides. Its distribution is determined by $V_{23}|V_{02} \sim F_{23}(x; V_{02}) = \mathcal{LS}^{-1}[\psi_{23}(\cdot\,; V_{02})](x)$, where $\psi_{23}(t; V_{02}) = \exp\big(-V_{02}\psi_2^{-1}(\psi_3(t))\big)$. To summarize, the tree-like (or cascading) dependence structure of a HAC stems from the distribution of the random variables V_0, V_{01}, V_{02} and V_{23}. Furthermore, note that the HAC (2) is simply survival copula [4, p. 36] of the random vector

$$\left(\frac{E_1}{V_0}, \frac{E_2}{V_{01}}, \frac{E_3}{V_{01}}, \frac{E_4}{V_{02}}, \frac{E_5}{V_{02}}, \frac{E_6}{V_{23}}, \frac{E_7}{V_{23}}, \frac{E_8}{V_{23}}\right).$$

Indeed, E_1/V_0 has survival function ψ_0, E_j/V_{01} has survival function ψ_1, $j \in \{2, 3\}$, E_j/V_{02} has survival function ψ_2, $j \in \{4, 5\}$, and E_j/V_{23} has survival function ψ_3, $j \in \{6, 7, 8\}$. To check the last, e.g., note that

$$\begin{aligned}
\mathbb{P}(E_j/V_{23} > t) &= \mathbb{P}(E_j > tV_{23}) = \mathbb{E}[\mathbf{1}_{E_j > tV_{23}}] = \mathbb{E}[\mathbb{E}[\mathbf{1}_{E_j > tV_{23}}|V_{23}]] = \mathbb{E}[\exp(-V_{23}t)] \\
&= \mathbb{E}[\mathbb{E}[\exp(-V_{23}t)|V_{02}]] = \mathbb{E}\big[\exp\big(-V_{02}\psi_2^{-1}(\psi_3(t))\big)\big] \\
&= \mathbb{E}\big[\mathbb{E}\big[\exp\big(-V_{02}\psi_2^{-1}(\psi_3(t))\big)|V_0\big]\big] = \mathbb{E}\Big[\exp\Big(-V_0\psi_0^{-1}\big(\psi_2\big(\psi_2^{-1}(\psi_3(t))\big)\big)\Big)\Big] \\
&= \mathbb{E}\big[\exp\big(-V_0\psi_0^{-1}(\psi_3(t))\big)\big] = \psi_0\big(\psi_0^{-1}(\psi_3(t))\big) = \psi_3(t).
\end{aligned}$$

3 Transformation using the Kendall distribution function

First, consider the sector copula C_3 and note that $(U_6, U_7, U_8) \sim C_3$. After having estimated the parameter(s) of C_3 (see [7] for some available AC estimators), one can let the vector (U_6, U_7, U_8) collapse to a single component $U_{K_{3,3}}$ in such a way that the parameters of the dependence structure of $(U_1, \ldots, U_5, U_{K_{3,3}})$ can be estimated. Due to the popularity of the Kendall distribution function, one might be tempted to choose $U_{K_{3,3}}$ as $K_{3,3}\big(\psi_3((E_6 + E_7 + E_8)/V_{23})\big)$, where $K_{k,l}$ denotes the Kendall distribution function in k dimensions based on ψ_l and the term $\psi_3((E_6 + E_7 + E_8)/V_{23})$ is the Kendall transformation of the vector (U_6, U_7, U_8), i.e., $\psi_3((E_6+E_7+E_8)/V_{23}) = \psi_3(\psi_3^{-1}(\psi_3(E_6/V_{23})) + \psi_3^{-1}(\psi_3(E_7/V_{23})) + \psi_3^{-1}(\psi_3(E_8/V_{23}))) = C_3(U_6, U_7, U_8)$. This approach is suggested by [14]. We will now show that $(U_1, \ldots, U_5, U_{K_{3,3}})$ does in general *not* have the copula

$$C_0\big(u_1, C_1(u_2, u_3), C_2(u_4, u_5, u_6)\big) \tag{4}$$

as distribution function. It follows that this estimation procedure is inconsistent for the parameters that correspond to all but the innermost sector copulas. This fact is supported empirically by the

simulation studies in [2, 14] and has also been suspected in [3]. In what follows, we derive the copula of $(U_1, \ldots, U_5, U_{K_{3,3}})$. Furthermore, we suggest an approach how to collapse (U_6, U_7, U_8) in such a way that we obtain a random vector following (4).

For $u_1, \ldots, u_6 \in [0,1]$, assume $x_1 = \psi_0^{-1}(u_1), x_2 = \psi_1^{-1}(u_2), x_3 = \psi_1^{-1}(u_3), x_4 = \psi_2^{-1}(u_4), x_5 = \psi_2^{-1}(u_5)$ and $x_6 = \psi_3^{-1}(K_{3,3}^{-1}(u_6))$. By stepwise conditioning on V_0, V_{01}, V_{02} and V_{23}, one can find that

$$\begin{aligned}
&\mathbb{P}\bigg(\frac{E_1}{V_0} > x_1, \frac{E_2}{V_{01}} > x_2, \frac{E_3}{V_{01}} > x_3, \frac{E_4}{V_{02}} > x_4, \frac{E_5}{V_{02}} > x_5, \frac{E_6+E_7+E_8}{V_{23}} > x_6\bigg) \\
&= \int_0^\infty \exp(-v_0 x_1)\bigg(\int_0^\infty \exp(-v_{01}(x_2+x_3))\, dF_{01}(v_{01}; v_0) \int_0^\infty \exp(-v_{02}(x_4+x_5)) \\
&\cdot \int_0^\infty \mathbb{P}(E_6+E_7+E_8 > V_{23}x_6 | V_{23} = v_{23})\, dF_{23}(v_{23}; v_{02}) dF_{02}(v_{02}; v_0)\bigg)\, dF_0(v_0). \qquad (5)
\end{aligned}$$

Taking into account the fact that the sum of n independent Exp(1) distributed random variables follows an Erlang distribution with survival function $\bar{F}_{Erl,n}(x) = \exp(-x)\sum_{k=0}^{n-1} x^k/k!$, $x \in [0,\infty)$, we obtain

$$\begin{aligned}
&\int_0^\infty \mathbb{P}(E_6+E_7+E_8 > V_{23}x_6 | V_{23} = v_{23})\, dF_{23}(v_{23}; v_{02}) = \int_0^\infty \mathbb{P}(E_6+E_7+E_8 > v_{23}x_6)\, dF_{23}(v_{23}; v_{02}) \\
&= \sum_{k=0}^2 \frac{(-x_6)^k}{k!} \int_0^\infty (-v_{23})^k \exp(-v_{23}x_6)\, dF_{23}(v_{23}; v_{02}) = \sum_{k=0}^2 \frac{(-x_6)^k}{k!} \int_0^\infty \frac{d^k}{dx_6^k} \exp(-v_{23}x_6)\, dF_{23}(v_{23}; v_{02}) \\
&= \sum_{k=0}^2 \frac{(-x_6)^k}{k!} \frac{d^k}{dx_6^k} \mathbb{E}[\exp(-V_{23}x_6) | V_{02} = v_{02}].
\end{aligned}$$

Using linearity, (5) therefore equals

$$\begin{aligned}
&\sum_{k=0}^2 \frac{(-x_6)^k}{k!} \frac{d^k}{dx_6^k} \int_0^\infty \exp(-v_0 x_1)\bigg(\int_0^\infty \exp(-v_{01}(x_2+x_3))\, dF_{01}(v_{01}; v_0) \\
&\cdot \int_0^\infty \exp(-v_{02}(x_4+x_5)) \mathbb{E}[\exp(-V_{23}x_6) | V_{02} = v_{02}]\, dF_{02}(v_{02}; v_0)\bigg)\, dF_0(v_0) \\
&= \sum_{k=0}^2 \frac{(-x_6)^k}{k!} \frac{d^k}{dx_6^k} \mathbb{E}\Big[\exp(-V_0 x_1)\mathbb{E}[\exp(-V_{01}(x_2+x_3)) | V_0] \\
&\qquad \cdot \mathbb{E}\Big[\exp(-V_{02}(x_4+x_5))\mathbb{E}[\exp(-V_{23}x_6)|V_{02}]\Big|V_0\Big]\Big] \\
&= \sum_{k=0}^2 \frac{(-x_6)^k}{k!} \frac{d^k}{dx_6^k} \mathbb{E}\Big[\exp(-V_0 x_1)\mathbb{E}[\exp(-V_{01}(x_2+x_3)) | V_0] \\
&\qquad \cdot \mathbb{E}\Big[\exp(-V_{02}(x_4+x_5)) \exp\big(-V_{02}\psi_2^{-1}(\psi_3(x_6))\big)\Big|V_0\Big]\Big] \\
&= \sum_{k=0}^2 \frac{(-x_6)^k}{k!} \frac{d^k}{dx_6^k} \mathbb{E}\Big[\exp(-V_0 x_1) \exp\big(-V_0\psi_0^{-1}(\psi_1(x_2+x_3))\big) \\
&\qquad \cdot \exp\Big(-V_0\psi_0^{-1}\big(\psi_2\big(x_4+x_5+\psi_2^{-1}(\psi_3(x_6))\big)\big)\Big)\Big] \\
&= \sum_{k=0}^2 \frac{(-x_6)^k}{k!} \frac{d^k}{dx_6^k} \psi_0\Big(x_1 + \psi_0^{-1}(\psi_1(x_2+x_3)) + \psi_0^{-1}\big(\psi_2\big(x_4+x_5+\psi_2^{-1}(\psi_3(x_6))\big)\big)\Big),
\end{aligned}$$

so that the copula of $(U_1, \ldots, U_5, U_{K_{3,3}})$ at $(u_1, \ldots, u_6)$ is not (4) but

$$\sum_{k=0}^2 \frac{\big(-\psi_3^{-1}(K_{3,3}^{-1}(u_6))\big)^k}{k!} \frac{d^k}{dx_6^k} C_0\big(u_1, C_1(u_2,u_3), C_2(u_4,u_5,\psi_3(x_6))\big)\Big|_{x_6=\psi_3^{-1}(K_{3,3}^{-1}(u_6))}. \qquad (6)$$

Although the implications of the representation (6) are unclear at the moment, we can interpret (6) as the Taylor polynomial of order two of $x_6 \mapsto C_0\big(u_1, C_1(u_2,u_3), C_2(u_4,u_5,\psi_3(x_6))\big)$ about $\psi_3^{-1}(K_{3,3}^{-1}(u_6))$ evaluated at zero. If the dimension of the sector copula C_3 converges to infinity, we get (with the same argument) the Taylor polynomial of order infinity and hence the function itself evaluated at 0, which is $C_0(u_1, C_1(u_2,u_3), C_2(u_4,u_5,\psi_3(x_6)))$ at $x_6 = 0$ and thus $C_0(u_1, C_1(u_2,u_3), C_2(u_4,u_5))$, which is the marginal copula of C for $u_6 = u_7 = u_8 = 1$.

4 Transformation using the diagonal of the Archimedean copula

As becomes directly clear from (2) or its tree representation in Figure 1, if we build marginal copulas of (2) for two of the components with index in $\{6,7,8\}$ being 1, we obtain (4). This means that after estimation of the parameters of C_2 one could (e.g., randomly) choose one of the components with index in $\{6,7,8\}$ for continuing with the parameter estimation on higher levels in (4). On the one hand, an advantage would be that the estimation error for the parameter of C_3 is not inherited. On the other hand, choosing just one component means reducing the available information. In what follows, we therefore describe a complementary procedure that is based on the copula diagonal of C_3 and shares the properties that are opposite to those mentioned above.

For constructing a random vector distributed according to (4), one can consider

$$U_{\delta_{3,3}} = \delta_{3,3}(\max\{U_6, U_7, U_8\}), \tag{7}$$

where $\delta_{k,l}(t) := \psi_l(k\psi_l^{-1}(t))$ denotes the diagonal in dimension k of the AC generated by ψ_l [7]. Note that $\mathbb{P}(U_{\delta_{3,3}} \le x) = \mathbb{P}(\max\{U_6,U_7,U_8\} \le \psi_3(\psi_3^{-1}(x)/3)) = \mathbb{P}(U_6 \le \psi_3(\psi_3^{-1}(x)/3)), U_7 \le \psi_3(\psi_3^{-1}(x)/3)), U_8 \le \psi_3(\psi_3^{-1}(x)/3))) = C_3(\psi_3(\psi_3^{-1}(x)/3),\, \psi_3(\psi_3^{-1}(x)/3), \psi_3(\psi_3^{-1}(x)/3)) = x$, provided C_3 is an AC with generator ψ_3, hence $U_{\delta_{3,3}} \sim \mathrm{U}[0,1]$, where $\mathrm{U}[0,1]$ denotes the univariate uniform distribution on $[0,1]$. Since the maximum is distributed according to the copula diagonal, see that $\delta_{3,3}^{-1}(U) \sim \delta_{3,3}$ for any $U \sim \mathrm{U}[0,1]$, one effectively replaces the copula C_3 by its diagonal. Assume $x_1,\dots,x_5$ as above and $x_6 = \psi_3^{-1}(u_6)$, and note that $\psi_3^{-1}(U_{\delta_{3,3}}) = 3\big(\psi_3^{-1}\big(\max\{\psi_3(\frac{E_6}{V_{23}}), \psi_3(\frac{E_7}{V_{23}}), \psi_3(\frac{E_8}{V_{23}})\}\big)\big) = 3\min\{\frac{E_6}{V_{23}}, \frac{E_7}{V_{23}}, \frac{E_8}{V_{23}}\}$. Then the resulting survival function can be computed as

$$\begin{aligned}
&\mathbb{P}\bigg(\frac{E_1}{V_0} > x_1, \frac{E_2}{V_{01}} > x_2, \frac{E_3}{V_{01}} > x_3, \frac{E_4}{V_{02}} > x_4, \frac{E_5}{V_{02}} > x_5, \frac{3\min\{E_6,E_7,E_8\}}{V_{23}} > x_6\bigg)\\
&= \int_0^\infty \exp(-v_0x_1)\bigg(\int_0^\infty \exp(-v_{01}(x_2+x_3))\,dF_{01}(v_{01};v_0)\int_0^\infty \exp(-v_{02}(x_4+x_5))\\
&\qquad\qquad \cdot \int_0^\infty \exp\Big(-v_{23}\Big(\frac{x_6}{3}+\frac{x_6}{3}+\frac{x_6}{3}\Big)\Big)\,dF_{23}(v_{23};v_{02})\,dF_{02}(v_{02};v_0)\bigg)\,dF_0(v_0)\\
&= \mathbb{E}\bigg[\exp(-V_0x_1)\mathbb{E}[\exp(-V_{01}(x_2+x_3))|V_0]\\
&\qquad \cdot \mathbb{E}\bigg[\exp(-V_{02}(x_4+x_5))\mathbb{E}\bigg[\exp\Big(-V_{23}\Big(\frac{x_6}{3}+\frac{x_6}{3}+\frac{x_6}{3}\Big)\Big)\Big|V_{02}\bigg]\Big|V_0\bigg]\bigg]\\
&= \mathbb{E}\Big[\exp(-V_0x_1)\mathbb{E}[\exp(-V_{01}(x_2+x_3))|V_0]\mathbb{E}\big[\exp(-V_{02}(x_4+x_5))\exp\big(-V_{02}\psi_2^{-1}(\psi_3(x_6))\big)\big|V_0\big]\Big]\\
&= \mathbb{E}\Big[\exp(-V_0x_1)\mathbb{E}[\exp(-V_{01}(x_2+x_3))|V_0]\mathbb{E}\Big[\exp\big(-V_{02}\big(x_4+x_5+\psi_2^{-1}(\psi_3(x_6))\big)\big)\Big|V_0\Big]\Big]\\
&= \mathbb{E}\Big[\exp(-V_0x_1)\exp\big(-V_0\psi_0^{-1}(\psi_1(x_2+x_3))\big)\exp\Big(-V_0\psi_0^{-1}\big(\psi_2\big(x_4+x_5+\psi_2^{-1}(\psi_3(x_6))\big)\big)\Big)\Big]\\
&= \mathbb{E}\Big[\exp\Big(-V_0\Big(x_1+\psi_0^{-1}(\psi_1(x_2+x_3))+\psi_0^{-1}\big(\psi_2\big(x_4+x_5+\psi_2^{-1}(\psi_3(x_6))\big)\big)\Big)\Big)\Big]\\
&= \psi_0\Big(x_1+\psi_0^{-1}(\psi_1(x_2+x_3))+\psi_0^{-1}\big(\psi_2\big(x_4+x_5+\psi_2^{-1}(\psi_3(x_6))\big)\big)\Big),
\end{aligned}$$

so that the copula of $(U_1,\dots,U_5,U_{\delta_{3,3}})$ at $(u_1,\dots,u_6)$ is indeed the HAC (4). We can therefore continue in the same fashion with estimating and further collapsing the sectors on the innermost nesting levels of this copula, i.e., in the same fashion as in the multi-stage procedure described in [14].

5 Conclusion

To summarize, when using the transformation based on the Kendall distribution function, the copula of $(U_4, U_5, U_{K_{3,3}})$ is generally not C_2. Hence, if we wrongly assume that $(U_4, U_5, U_{K_{3,3}}) \sim C_2$, then even if a consistent estimator for the parameters of the copula C_2 is used, the resulting estimator is in general inconsistent for the parameter(s) of C_2. In contrast, if the transformation based on the diagonal of the AC is used (or any component with index in {6, 7, 8}), it is assured that the sector copulas in higher nesting levels of the estimated HAC are the distribution functions of the corresponding random vectors, i.e., $(U_4, U_5, U_{\delta_{3,3}}) \sim C_2$ and $(U_1, U_{\delta_{2,1}}, U_{\delta_{3,2}}) \sim C_0$. Using any consistent estimator for the parameters of an AC, the resulting estimator for the parameters of the HAC is consistent.

Acknowledgements

The work of Jan Górecki was funded by the project SGS/21/2014 – Advanced methods for knowledge discovery from data and their application in expert systems, Czech Republic. The work of Martin Holeňa was funded by the Czech Science Foundation (GA ČR) grant 13-17187S.

References

[1] Grecki, J., Holea, M.: An alternative approach to the structure determination of hierarchical Archimedean copulas. In: *Proceedings of the 31st International Conference on Mathematical Methods in Economics (MME 2013)*, Jihlava, 2013, 201–206.

[2] Grecki, J., Holea, M.: Structure determination and estimation of hierarchical Archimedean copulas based on Kendall correlation matrix. In: *Proceedings of the 2nd Workshop on New Frontiers in Mining Complex Patterns (NFMCP 2013)*, 2013, 2–16.

[3] Hering, C.: *Estimation Techniques and Goodness-of-fit Tests for Certain Copula Classes in Large Dimensions*. PhD thesis, Ulm University, 2011.

[4] Hofert, M.: *Sampling Nested Archimedean Copulas with Applications to CDO Pricing*. Suedwestdeutscher Verlag fuer Hochschulschriften, 2010.

[5] M. Hofert. Efficiently sampling nested Archimedean copulas. *Computational Statistics and Data Analysis*, 55(1):57–70, 2011.

[6] Hofert, M.: A stochastic representation and sampling algorithm for nested Archimedean copulas. *Journal of Statistical Computation and Simulation* **82**, 9 (2012), 1239–1255.

[7] Hofert, M., Mächler, M., McNeil, A. J.: Archimedean copulas in high dimensions: Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique* **154**, 1 (2013), 25–63.

[8] Hofert, M. Scherer, M.: CDO pricing with nested Archimedean copulas. *Quantitative Finance* **11**, 5 (2011), 775–787.

[9] Joe, H.: *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, 1997.

[10] Kimberling, C. H.: A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicae* **10** (1974), 152–164.

[11] McNeil, A. J.: Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation* **78**, 6 (2008), 567–581.

[12] McNeil, A. J., Nešlehov, J.: Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *The Annals of Statistics* **37** (2009), 3059–3097.

[13] Nelsen, R.: *An Introduction to Copulas*. 2nd edition, Springer, 2006.

[14] Okhrin, O., Okhrin, Y., Schmid, W.: On the structure and estimation of hierarchical archimedean copulas. *Journal of Econometrics* **173**, 2 (2013), 189–204.

[15] Sklar, A.: Fonctions de rpartition a n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **8** (1959), 229–231.

The trends in the income distributions in the EU-27 countries: Measuring the differences

Vladimír Hajko [1], Naďa Birčiaková [2], Jana Stávková [3]

Abstract. This article investigates the trends in the income distributions (measured in the PPS) in the EU-27 countries during the period 2005-2013. This article focuses on the income development from two perspectives. First, it examines the influence of the crisis period on the specific income groups. The results show the income development was negatively influenced in all examined groups and categories. Relatively highest impact (compared to the previous trend) can be observed for the income of the poorest households, which basically stopped growing. Second, this article examines whether there is an observable income convergence in the specific country groups, both unconditional and dependent on the income level group. The results show the half-life of convergence is around 12.4 years for the median income category in the EU27 country group. The convergence rate is slightly faster with increasing income category. However, much faster convergence rates of around 4 years can be observed in the relatively poor (Group 01) and relatively rich (Group 03) country groups than in the middle group.

Keywords: income, development, distribution, crisis

JEL classification: D31
AMS classification: 62J05

1 Introduction

The disposable income is one of the most significant measures related to the household behavior. The economic logic dictates the rate of satisfaction of the individual needs is primarily limited by the budget constraint. In consequence, the income development is among the most important topics in the social policy investigation.

Multiple social theories (including predominantly economics and sociology) point out the income is in very close relation to the standard of living or its broader concept of quality of life. In consequence the income and income distribution is among the main focus points of both the economists and the government policies.

Despite this fact, much remains to be investigated regarding the income development - see for instance [4] for the discussion regarding the income inequality, or [5] for the investigation of the link between economic performance, institutional quality and income distribution.

The international comparison of both the income development across the countries, and income distribution within the countries are among the most popular topic in social policy debates. Some authors argue that the increasing income inequality might be reflected in the negative effects for the society, or even negatively influence the economic growth [2, 13].

On the other hand, some authors argued that the higher income inequality might be positively associated with growth [6, 15].

The conflicting empirical evidence (and consequent theoretical implications) for the actual impacts of the inequality itself has yet to be resolved. In this article we focus on a phenomenon that seems to be observable in many countries, i.e. differentiated development of income across the income groups as well

[1] Mendel University, Faculty of Business and Economics (FBE), Zemědělská 1, Brno, Czech Republic, vladimir.hajko@mendelu.cz
[2] Mendel University, FBE, nada.birciakova@mendelu.cz
[3] Mendel University, FBE, jana.stavkova@mendelu.cz

as the differentiated responses to the economic downturns (see for instance [7]).
These differentiated responses might be interesting both from the theoretical viewpoint (indicating the underlying differences in the data generating processes for different groups) and from the practical policy application.
Somewhat different point of view regarding the income development is provided in the literature dealing with the income convergence ([1] and [8] provide extensive discussion).
This rather popular branch of research started as a response to the articles by [14] and [10]. In the context of the macroeconomic growth theory the most prominent responses (often viewed as seminal in the subsequent empirical literature) were [3] and [11]. Here the main interest is not as much the individual income, but rather aggregate values - typically the average income per capita - in relation to the macroeconomic development of individual countries towards the steady state.
Apart from the seminal articles, multiple subsequent studies have investigated the empirical convergence patterns. The general notion of the conditional β-convergence seems to be sufficiently robust even for global samples. But the global samples are generally unable to provide the evidence for the unconditional convergence.
The convergence debate however seems to agree on the existence of convergence patterns within the groups of the developed and within the groups of developing countries (see e.g. [9], finding the evidence for the convergence among the African countries). However, it seems a divergence between these two groups can be observed.
The income convergence is also the interest of the investigation of the regional characteristics within a country. An example of such intra-country approach (on very long sample) is [12], finding strong convergence pattern among twenty-four Swedish counties.
This article aims to combine the two aforementioned distinct approaches, with the focus primarily on the income situation of households. If there are differentiated responses present in the selected quantile groups regarding the income development in the EU countries, it will indicate that some of the groups are able to better respond to the unfavorable economic conditions. In the low income groups, this type of observation might be caused by the social programs substantial enough to facilitate unchanged income pattern for the recipients of the social programs (for instance at the cost of increased government deficits). The alternative view would suggest that the income situation is largely influenced by the human capital. In that case, the poorest households and unskilled workers would be the first to be negatively influenced by the crisis.
The typical expectation for the high income group, would be that the higher income opportunities and potentially higher levels of human capital in this group might allow to adjust to the economic downturn better than other groups. Therefore we might expect lower reductions in the standard of living.
In this article we aim to find out which income category was influenced the most. Furthermore, as the European Union suffers from the significant heterogeneity of the income levels in its member countries (recall the Convergence objective of the EU Structural Funds), we investigate if the individual quantile groups are converging (both in general and within a selected country group), and if yes, whether there are differences in the rate of convergence (this can be also viewed as the investigation whether the EU Convergence objective is being fulfilled).

2 Data

The data come from the European Union Statistics on Income and Living Conditions survey (EU SILC). The aggregated results of the regular survey are available from the Eurostat. The income was measured in the units of purchasing power standard (PPS) to avoid the influence of different cost levels in the individual countries. The income variable is measured as the disposable income of the households, i.e. a representation of personal income (I_p) (all income from work, such as wage or income from self-employment + income from investment and property) adjusted with social transfers (Tr) and taxes (T). The simple representation of the disposable income (DI) is therefore $DI = I_p + Tr - T$. The income measurement excludes imputed rent and non-monetary income components (e.g. value of goods produced for own consumption). Since the measurement is carried out for households[1] of variable size, the income measure is equivalized. The Eurostat applies equivalization factors that assigns a weight of 1.0 to the first person aged 14 or more, a weight of 0.5 to other persons aged 14 or more and a weight of 0.3 to persons aged 0-13. The equivalized income of the household with three adults and two children (below 13) would

[1]The household denotes a person living alone or a group of people who live together in the same private dwelling and share expenditures, including the joint provision of the essentials of living.

therefore be calculated as $DI_{eqv} = DI * (1 + 0.5 * 2 + 0.3 * 2)^{-1} = \frac{DI}{2.6}$. Note that in the further text we will denote the equivalized income simply as DI.

The values of household income selected for the investigation of the articles represent the 5th, 20th, 50th, 80th and 95th percentile of the equivalized income distribution in the given year.

3 Methodology

In order to distinguish the structural differences between "poor" and "rich" countries, i.e. to capture the underlying distributional heterogeneity, we divided the countries into three groups, based on the comparison of the average economic level (measured by the GDP per capita in PPS (data available from Eurostat Table tec00114).

Based on this selection, the first group includes (the order of countries is ascending with the economic level): Bulgaria, Romania, Croatia, Latvia, Hungary, Poland, Estonia and Lithuania (with economic levels ranging from 47 to 72 percent of EU28 average in 2012).

The second group includes Greece, Portugal, Slovakia, Czech Republic, Slovenia, Malta, Cyprus, Spain, Italy, United Kingdom, France and Finland (with economic levels ranging from 75 to 115 percent of EU28 average in 2012).

The third group includes Belgium, Germany, Denmark, Sweden, Netherlands, Ireland and Austria (with economic levels ranging from 120 to 130 percent of EU28 average in 2012).

The single remaining country of Luxembourg is a clear outlier with the economic level of 263 percent of EU28 average in 2012 and was therefore omitted from the working sample.

Figure 1 shows that even this simple distinction of groups is sufficiently helpful in the elimination of the double peaks in the distributions of income across the individual country groups.

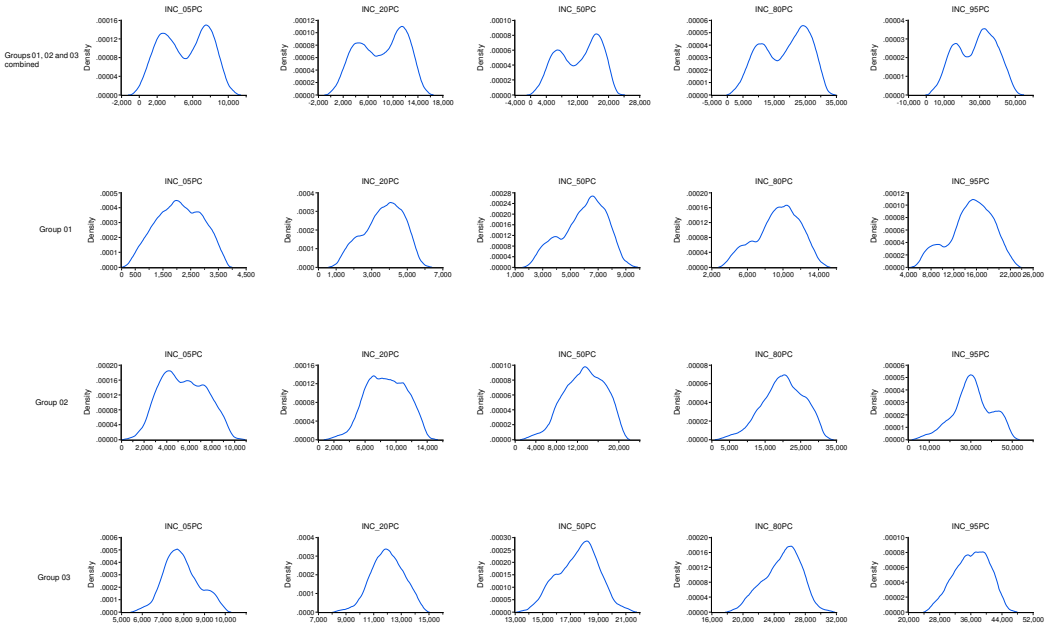


Figure 1 Estimated Kernel densities, by groups and by income quantiles, EU-27 countries

3.1 Model specification

We test whether there was an observable difference in the trends that can be attributed to the period of crisis (we defined the beginning of the crisis in 2008). In other words, we investigate the significance of the dummy period representing the period 2008-2012 on the intercept and trend. The estimated relationship is:

$$DI_{i,t} = \alpha_i + \beta T_t + \delta D1_t + \gamma D1_t T_t + \epsilon_{i,t} \quad (1)$$

The test hypothesis can be described as restricting the parameters δ and γ in (1).

In order to investigate the convergence patterns, we estimate the following (unconditional) convergence model:

$$\Delta log\left(DI_{i,t}\right) = \alpha + \beta log\left(DI_{i,t-1}\right) + \varepsilon_{i,t} \tag{2}$$

Note that the conditional convergence would assume multiple steady states conditional on the country-specific characteristics μ_i replacing the α in the previous equation. This would assign country-specific effect to each cross-sectional unit. We are estimating (2) for the whole sample (of EU countries except Luxembourg) and for three subgroups described above in section 3.

4 Results

Table 1 shows the results of the estimation based on (1). The 5th percentile of households (by income) show the most significant change in the trend, with more than 100% values of the original trend coefficients. Surprisingly enough, the highest change can be recorded for the 5th percentile income category in the Group 03. The remaining income levels recorded on average -79.7% change in the trend coefficient in the crisis period.
A common pattern across the country groups is that the impact of the crisis period (in terms of % change of the trend component in the income) is decreasing with increasing economic level (i.e. with each subsequent country group). This holds for all examined income categories except the lowest one (5th percentile). The lowest impact of the crisis period was recorded for 80th percentile income category in all groups and the relatively smallest impact can be observed in the 50th and 80th percentile income categories in Group 03.
The β-convergence results are provided in Table 2. Instead of the β coefficients, the table reports two typical measures of the convergence speed - the so called implied speed of convergence (typically denoted λ)[2] and the half-life of convergence (H)[3].
The half-lives of the convergence for all income categories are rather similar for the EU-27 countries. As is typical, a faster convergence rate can be observed if we introduce the conditional convergence. In this case we examine the convergence pattern by country groups 01 through 03. The half-lives for the Group 01 and Group 03 lie in the vicinity of 4 years, while the Group 02 (admittedly the most heterogeneous one of the three) shows values in the vicinity of 8 years. A striking exception is the very slow convergence rate among the lowest income category in the Group 02. If we assume the lowest income category is the most likely recipient of the transfer payments, this probably indicates the significantly different responses of the social policy in the Group 02 countries.

5 Conclusion

This article investigated the income development for 5 income categories in the EU countries (excluding Luxembourg as obvious outlier) during the period 2005-2013. The EU countries were examined as three separate groups, divided by their income levels (division points between the categories were specified as 72% and 115% of EU28 income level in 2012) and the union of these three groups.
Specifically we have examined the change in the trend components of the income during the crisis period (beginning in 2008). The most significant changes can be observed in the lowest income category (5th percentile) in all examined country groups. Basically, the household income of the poorest households has stopped growing since the crisis began and in some cases is even in decline. However, it is clear that income development was negatively influenced in all examined groups and categories and there is no evidence that some income categories made a "profit" from the crisis.
We also investigated the β-convergence among the selected groups. The rate of convergence the Group 01 and 03 is rather similar, with half-lives (H) in the vicinity of 4 years. The convergence rate of the Group 02 is somewhat slower, with H around 8 years. However, if we take a look at the convergence per income category only, the convergence rate is slightly faster with each subsequent higher income category (ranging from $H = 12.6$ for the 05th percentile to $H = 11$ for the 95th percentile).

[2]Implied speed is given by: $\lambda = -\frac{ln(1+\hat{\beta})}{T}$, where T represents the time span used to measure the growth rates.

[3]The half life (H) can be derived from the expression $e^{-\lambda H} = 0.5$ and represents the approximate number of years it would take for half of the current disparities in the group to be eliminated.

		EU-27	Group 01	Group 02	Group 03
05th percentile					
	intercept	4464.5	1246.5	4786.2	7256.7
		(73.7)	(101.1)	(123.8)	(135.6)
	timetrend	266.4	311.2	277.1	218.6
		(56)	(72.9)	(95.9)	(105)
	D1	1001.3	1110.5	948.2	1028.8
		(144)	(175.2)	(250.6)	(283.6)
	D1*timetrend	−277.2	−320.5	−281.4	−246.2
		(60.8)	(77.7)	(104.5)	(115.7)
20th percentile					
	intercept	6981.5	2398.8	7584.6	10758.6
		(80.5)	(108.3)	(143.3)	(122.1)
	timetrend	435.2	431	434	439.2
		(61.2)	(78.2)	(111)	(94.6)
	D1	1179.9	1244	1119.1	1181.6
		(157.2)	(187.9)	(290.1)	(255.4)
	D1*timetrend	−347.7	−360	−344.9	−329.6
		(66.4)	(83.3)	(120.9)	(104.2)
50th percentile					
	intercept	10564.5	3960.8	11643.9	15615.3
		(125.2)	(203.4)	(204.4)	(216)
	timetrend	656	675.5	636.4	678.6
		(95.2)	(146.8)	(158.3)	(167.3)
	D1	1841.2	2262.1	1688.8	1574.5
		(244.6)	(352.8)	(413.9)	(451.8)
	D1*timetrend	−528.4	−618.8	−511.4	−446.8
		(103.2)	(156.5)	(172.6)	(184.3)
80th percentile					
	intercept	15665.1	6404.5	17618.5	21958.3
		(195.7)	(330.2)	(314.5)	(331.9)
	timetrend	969.3	998.9	842.1	1178.9
		(148.8)	(238.2)	(243.6)	(257.1)
	D1	2644.1	3392	2313.5	2295.7
		(382.4)	(572.5)	(636.7)	(694.2)
	D1*timetrend	−754.2	−898.1	−644.3	−770.8
		(161.4)	(254)	(265.4)	(283.3)
95th percentile					
	intercept	23842.1	10760.9	27220	31697.3
		(356.3)	(592.8)	(586)	(600.7)
	timetrend	1380.1	1355.6	1173.1	1763.9
		(270.8)	(427.7)	(453.9)	(465.2)
	D1	4167.9	5114.6	3689.7	3747.1
		(696.1)	(1028.1)	(1186.5)	(1256.4)
	D1*timetrend	−1113.3	−1268	−910	−1240.5
		(293.8)	(456)	(494.6)	(512.6)

Coefficient values, std. errors in parentheses

Table 1 Trend break results

		EU-27	Group 01	Group 02	Group 03
05th percentile	Implied speed	0.055	0.158	0.032	0.216
	Half-life	12.6	4.4	21.7	3.2
20th percentile	Implied speed	0.051	0.124	0.072	0.194
	Half-life	13.6	5.6	9.6	3.6
50th percentile	Implied speed	0.056	0.164	0.082	0.246
	Half-life	12.4	4.2	8.5	2.8
80th percentile	Implied speed	0.059	0.167	0.083	0.231
	Half-life	11.7	4.2	8.4	3
95th percentile	Implied speed	0.063	0.181	0.088	0.185
	Half-life	11	3.8	7.9	3.7

Table 2 Convergence estimates

References

[1] Maria Abreu and Henri L. F. de Groot, *A meta-analysis of β-convergence: The legendary 2%*, Journal of Economic Surveys **19(3)** (2005), 389–420.

[2] A. Alesina and D. Rodrik, *Distributive politics and economic growth*, Quarterly Journal of Economics **109** (1994), 465–490.

[3] R.J. Barro and X.X. Sala-i Martin, *Convergence*, Journal of Political Economy **100** (1992), 223–251.

[4] Roland Benabou, *Unequal societies: Income distribution and the social contract*, The American Economic Review **90** (2000), 96–129.

[5] Alberto Chong and Csar Calderón, *Institutional quality and income distribution*, Economic Development and Cultural Change **48** (2000), 761–786.

[6] W. Easterly and S. Rebelo, *Fiscal policy and economic growth: an empirical investigation*, Journal of Monetary Economics **32** (1993), 417–455.

[7] B. Habib, A. Narayan, S. Olivieri, and C. Sanchez, *The impact of the financial crisis on poverty and income distribution: Insights from simulations in selected countries*, World Bank Economic Premise Note Series, March 2010, No. 7.

[8] N. Islam, *What have we learnt from the convergence debate?*, Journal of Economic Surveys **17** (2003), 309–362.

[9] Faiza A. Khan, *Economic convergence in the african continent: Closing the gap*, South African Journal of Economics **1** (2013), 1–17.

[10] R.E. Lucas, *On the mechanics of economic development*, Journal of Monetary Economics **22** (1988), 3–42.

[11] N. Mankiw, D. Romer, and D. Weil, *A contribution to the empirics of economic growth*, Quarterly Journal of Economics **107** (1992), 407–437.

[12] Joakim Persson, *Convergence across the swedish counties, 1911-1993*, European Economic Review **41** (1997), 1835–1852.

[13] T. Persson and G. Tabellini, *Is inequality harmful for growth: theory and evidence*, American Economic Review **84** (1994), 600–622.

[14] P.M. Romer, *Increasing returns and long-run growth*, Journal of Political Economy **94** (1986), 1002–1037.

[15] Kevin Sylwester, *Income inequality, education expenditures and growth*, Journal of Development Economics **63** (2000), 379–398.

Simulation of data for reforestation system

David Hampel[1], Jitka Janová[2]

Abstract. Only limited real data related to wood processing economy are available, because companies do not present their costs and profits in regular way. We need to extend obtained data set to cover overall situation in forestry management, and estimate revenue and cost functions. These are predefined according to formulation of underlying optimal control problem and it is necessary to estimate coefficients of these functions. First, we describe and realize simulation procedure for obtaining data illustrating plantations of various mixed forests in the Morava region in the Czech Republic. Second, we estimate and discuss parameters of functions suitable for subsequent optimization.

Keywords: biodiversity, economic profit, forestation, regression, simulation.

JEL classification: C53
AMS classification: 91B74

1 Introduction

Spruce forests, in most cases artificially planted, cover the majority of forest land in the Czech Republic. Spruce timber currently belongs to the most versatile used softwoods. It is processed by furniture and construction industry in large quantities, because it is easy-to-process (it is soft wood), tough, relatively strong and flexible. Spruce is an important raw material for paper production. A special feature is the so-called resonant spruce, which is indispensable in the production of musical instruments.

Despite the desirable properties of spruce timber, the problems with exhausted soil and lack of biodiversity in forests have escalated to intensive need for modifying the established forest structure. There are number of possible particular combination of mixed forests to involve, various rotation scenarios to choose from and number of different forestation strategies to adopt.

Desired changes in forest species structure can be achieved by concerning the appropriate methods of operations research, where we can distinguish the static and dynamic approach. These methods would provide the support for decision making. The change must be economically manageable, so the production and profit functions for particular forest types must be taken into account.

We assume the dynamic optimization techniques to be appropriate when identifying the optimal long run reforesting strategy for the forest owners in the Czech Republic. We propose to employ the optimal control for finding the optimal time path for the state variables – the areas forested by the particular forest types. We expect that for the infinite horizon problem with (current value) profit maximization criterion we obtain a steady state solution representing the optimal target forest structure and the optimal state and control paths determining the appropriate strategies of forestation. Basic description of this approach is given in [5] and further results are under submission process in the Central European journal of operations research currently.

We focus on the optimal forestation strategy from the point of view of the forest owners in the Drahanska highlands area (localized in the Morava region – the east half of the Czech Republic) with 80 800 ha of forest land given the current forest structure. In our study we focus on the five most common tree species which will leave us with total forest land considered 72 122 ha, see Tab. 1.

Solving of optimal control problem requires some parameters to be estimated. In our case, it is done with the comprising of simulation techniques. In the field of forestry, simulation techniques are used very

[1]Mendel University in Brno, Department of Statistics and Operational Analysis, Zemědělská 1, Brno, Czech Republic, qqhampel@mendelu.cz
[2]Mendel University in Brno, Department of Statistics and Operational Analysis, Zemědělská 1, Brno, Czech Republic, janova@mendelu.cz

i	1	2	3	4	5
Tree specie	Beech	Oak	Pine	Larch	Spruce
$x_i(0)$ [ha]	12140	6715	9942	6471	36854

Table 1 Tree species i with the current forested area $x_i(0)$

often – especially for investigating growth or another production characteristics of a forest. Some authors add ecological, environmental or social aspects to raw production functions of a forest, see for example [6]. Beside this economical aspects of forest planting are explored, see [9]. Ecological dimension engages also in primarily economical papers, see for example [13], [7] and [8], where enviro-economic aspects of particular mixed forests are discussed.

The aim of this paper is to simulate and verify forestation cost and revenue data suitable for further regression analysis designed for optimal control problem solution. The rest of the paper is organized as follows: section 2 introduces motivation of our research in exact terms, data and methodology, section 3 comprises results and the last section concludes.

2 Material and methods

This section include motivation of our work, which is important to understand the reasons why we deal with the presented problem. Further we discuss available real data. Finally, description of simulating procedure is given.

2.1 Motivation

The final goal of our work is to solve further dynamic decision problem: to find the optimal subsidy strategy for reaching the desired forest structure in the Czech Republic within the least time. In our model, the criterion of decision making is the economic profit from forests (including the income from selling the timber and obtaining the subsidies). To address the decision problem we introduce infinite horizon optimal control problem with free terminal points, which can be schematically described as

$$\max V = \int_0^\infty \Pi(t) e^{-\rho t} \mathrm{d}t, \tag{1}$$

where

$$\Pi(t) = \sum_{i=1}^{n} R_i(x_i(t), u_i(t)) - K_i(x_i(t), u_i(t)) \tag{2}$$

under appropriate constrains. Here, n denotes the number of tree species appropriate for the given region, $x_i(t)$ are the state variables representing the area of land forested by specie i in time t, ρ is discount rate. Further we denote total area reforested at time t by specie i (in hectare per year) $u_i(t)$, the current profit $\Pi(t)$, total revenues from selling the timber and subsidies $R_i(t)$ and total costs of growing and logging at time t $K_i(t)$. Note that we do not consider forestation of non-forest land in our model.

Optimal control techniques have been steadily employed in the forestry management research problems during the several last decades for analyzing the potential optimal strategies (for recent contributions mostly related to our problem see [2], [3] [1] or [4]). However, due to the complex structure of the dynamic decision problems in forestry, some theoretically oriented papers have appeared addressing the particular specifics of forestry decision problems [12].

We assume that

$$R_i(t) = G_i(x_i(t)) + \sigma_i u_i(t), \tag{3}$$

where G_i represents a known function of revenues from logging and $\sigma_i u_i(t)$ is the subsidy from increasing the area of specie i. In current stage of the problem solving we suppose constant subsidies. Note that only the broadleaved species are considered to be supported by subsidies.

For the estimation of the functions $G_i(x_i)$ and $K_i(x_i, u_i)$ we adapt the assumptions made in [2] and set

$$G_i = g_{i0} + g_{i1}x_i + \frac{1}{2}g_{i2}x_i^2, \tag{4}$$

$$K_i = k_i + a_i x_i + b_{i1}u_i + \frac{1}{2}b_{i2}u_i^2, \tag{5}$$

where the total cost function (5) is composed from the farming costs and reforestation costs. We assume $b_{i1} = b_{i2} = 0$ for the spruce type forest, which is supposed to be reduced. It means that decreasing the area of the forest type (after logging the current stand) is costless. To proceed with the solution of optimal control problem (1) it is necessary to estimate parameters of the functions (4) and (5).

2.2 Data

Only limited real data related to wood processing economy are available, because companies do not present their costs and revenues in a regular way. We have collected the original data on the local forest structure, land characteristics and technological-economical characteristics of forest growing, thinning and the timber selling.

The data were acquired from the state agencies (Czech Statistic Office, The Forest Management Institute) and from the Mendel University forest company (managing 10 000 ha of forest land in Morava region of the Czech Republic). Data have the form of unit cost or revenue tabulation mainly, which omits possible changes given by the size of stumped area.

Further, obtained data were discussed with experts from the fields of forestry and wood processing. Their opinions were useful for the verification of explanatory power of the data. Moreover, it brings increased accuracy of data for particular growing conditions.

2.3 Simulation study

According to datasets insufficient for immediate regression analysis, a simulation procedures were suggested to generate the missing data using the information available. For generating of the new samples we use common techniques stated for example in [10] or [11]. Simulation is provided simultaneously for five classes of single-type wood (spruce, oak, beech, larch, pine), where costs and revenues related to lodging and reforesting can be established using known data.

Initially, area of the forest is simulated. For costs determination, percentage of cutted down part of the forest is simulated. Costs of different tree species logging are not the same. Moreover, overall costs depend on soil quality distribution, which affects number of trees for cutting down and consequently cost of logging. The situation is further complicated by the fact, that a relief of the forest is not always flat. In some cases, it is necessary to use simple cableway to transport trees, what increases the cost of lodging.

We also capture limited working forces factor. According to this, price for lodging one cubic meter of timber cannot be constant, but increases: for small areas slightly, for very large areas significantly. The same situation we assume for the tree planting. Of course, for lodging this is true in short-time horizon only. Over the long term, heavy equipment may be deployed and it can cause constant or decreasing unit costs for increasing area of lodging. However, the clear-cut logging is not permitted everywhere (apparently, it is not in accordance with sustainable forests planting) so heavy equipment is not solution always.

Further we determine percentages of cutted down part of the spruce forest designated for reforestation by other tree species. This determines cost of reforestation, because reforestation by different trees is differently expensive. We assume that oaks, beeches, larches and pines are replaced by the same species. Costs related to forest cultivation, as pruning and weeds cutting, are incorporated too. Process of costs simulation can be expressed symbolically as

$$costs \sim f(area, soil\ quality, terrain\ profile, working\ power, reforestation).$$

Simulation of revenues uses previous ideas partially. Above them, distributions of prices related to different quality (and purpose) timber coming from different species are estimated from available historical

datasets. We use data from the years 2008–2012 which are relatively stable in mean. Estimated distributions are used for simulating of the one cubic meter timber revenues. Note that higher soil quality generally means higher timber quality. Process of revenues simulation we can express as

$$revenues \sim f(area, soil\ quality, prices\ of\ different\ quality\ timber).$$

Schematic illustration of both the costs and the revenues simulation is shown in Fig. 1. Costs and revenues are recalculated according to different life cycle of particular species, because this aspect cause further optimization significantly.

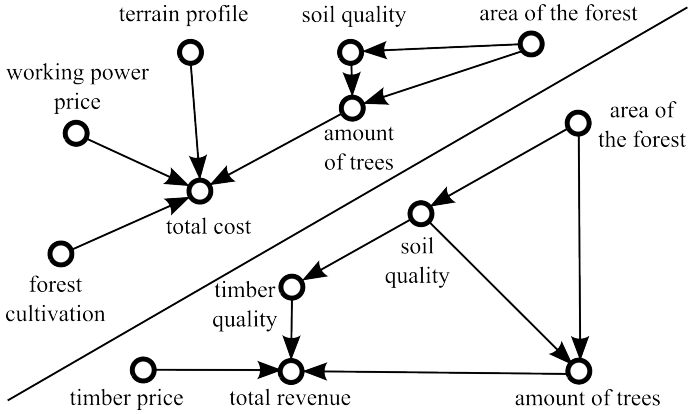


Figure 1 Scheme of costs and revenues simulation

3 Results and discussion

For all tree species we simulated 100, 1000, 10 000 and 100 000 cost and profit values gradually. These simulations were repeated one hundred times. For appropriate simulated values we estimate coefficients of cost and revenue functions (4) and (5). All parameters except constants are statistically significant at the significance level $\alpha = 0.05$. Coefficients of determination are always higher than 0.9, which point out high quality of particular regressions.

Now we will discuss stability of estimated parameters. As a model tree specie we will use oak; results are similar to other species. For each size of simulation we calculate average regression parameters estimate and average lower and upper limits of respective confidence interval. For all calculations, i.e. simulation and regression, we used computational system Matlab R2014a.

Results are shown in Fig. 2 and 3. The width of confidence interval diminishes with increasing size of simulation. Average value of estimated parameter is stable; in the case of the cost function (5) it holds for all simulation sizes, in the case of the revenue function (4) it is true from 1000 simulated values.

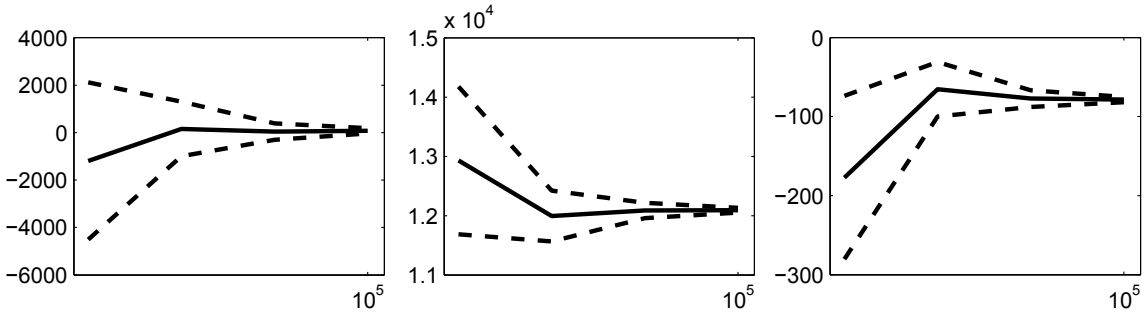


Figure 2 Average parameters of revenue function (4) (solid line) and confidence interval limits (dashed lines) for oak. On the x-axis is the size of simulation. Parameter g_{20} left graph, parameter g_{21} the middle graph, parameter g_{22} right graph.

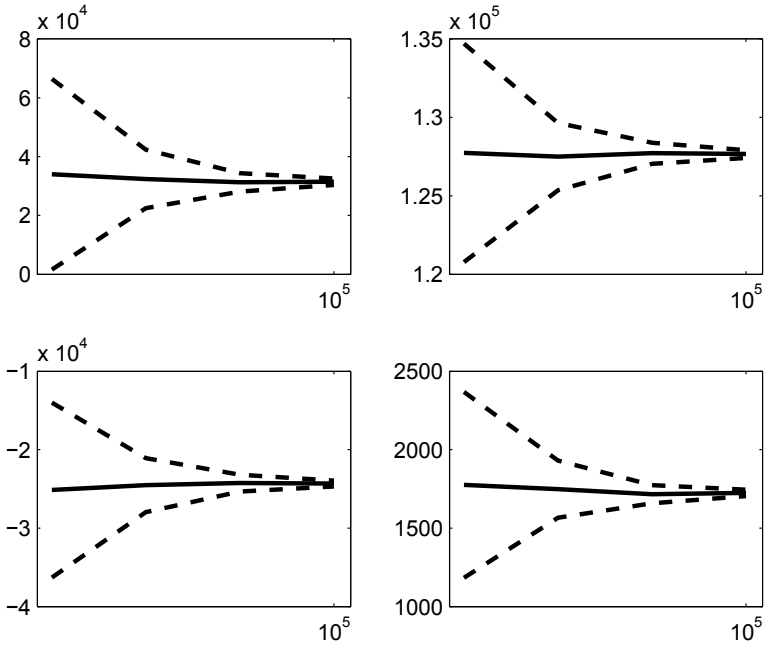


Figure 3 Average parameters of cost function (5) (solid line) and confidence interval limits (dashed lines) for oak. On the x-axis is the size of simulation. Parameter k_2 top left graph, parameter a_2 top right graph, parameter b_{21} bottom left graph, parameter b_{22} bottom right graph.

Further we discuss plausibility of estimated functions (4) and (5). In Fig. 4 we can see shape of these functions. We can see that cost function is greater than revenue function systematically. This finding confirms important role of subsidies in forestry. For the oak, it is possible to receive subsidies 105 025 CZK per hectare of newly planted forest currently. It is visible, that this subsidy allow to create appropriate profit, see Fig. 4 dashed line, where function (3) is above cost function.

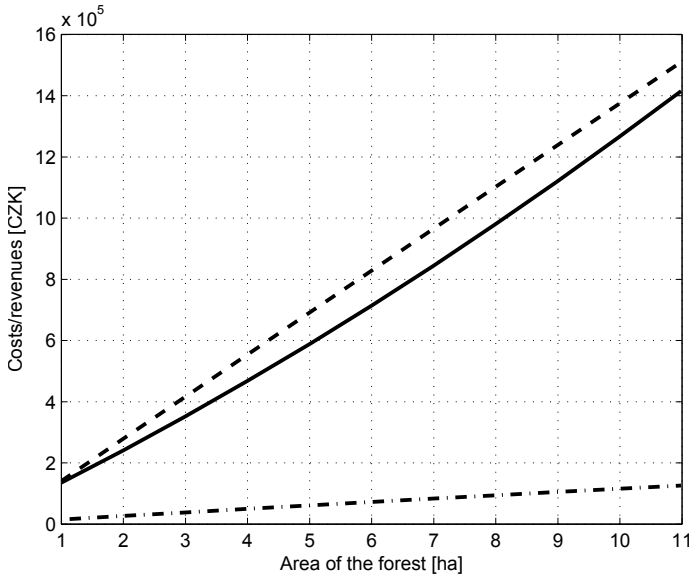


Figure 4 Cost function (5) (solid line), revenue function (4) (dot-and-dash line) and total revenue function including subsidies (3) (dashed line). In this graph holds $u_2 = 1.2x_2$.

4 Conclusions

We have developed the simulation procedure that enabled obtaining of the cost and revenue functions. These are required for solution of optimal control problem we are interesting in. Parameters estimated

for the cost and revenue functions seems to be stable, and we can conclude that 100 000 simulated values ensure results of the highest possible quality.

Further we ensure that the resulting functions are reliable representations of the real economical procedures in the forestry companies. The results obtained can be used as an indicator how the subsidies and other economic characteristics would drive the decision making of the forest owners in the Czech Republic. Our results correspond with the expert opinions of forestry managers and professionals.

Among further tasks we include the fact that apart from economic criterion we should mention other benefits from forest planting, that could enter the model – for example soil degradation. The next task is to investigate not only one-species forests, but forests with mixture of two or more species. Also we want to explore impacts of birch growing in the future.

Acknowledgements

Supported by the grant No. 13-25897S of the Czech Science Foundation.

References

[1] Bach, C. F.: Economic incentives for sustainable management: a small optimal control model for tropical forestry, *Ecological Economics* **30** (1999), 251–265.

[2] Caparros, A., Cedra, E., Ovando, P., and Campos, P.: Carbon sequestration with reforestations and biodiversity-scenic Values, *Environmental Resource Economics* **45** (2013), 49–72.

[3] Caparros, A., and Jacquemont, F.: Conflicts between biodiversity and carbon sequestration programs: economic and legal implications, *Ecological Economics* **46** (2013), 143–157.

[4] Cerda, E., and Martin-Barroso, D.: Optimal Control for Forest Management and Conservation Analysis in Dehesa Ecosystems, *European Journal of Operational Research* **227** (2013), 515–526

[5] Janová, J., and Kadlec, J.: Optimal Control for Forest Management in the Czech Republic. In: *Proceedings of the 3rd International Conference on Operations Research and Enterprise Systems*, Science and Technology Publications, Portugal, 2014, 466–470.

[6] Karjalainen, T., and Kellomki, S.: Simulation of forest and wood product carbon budget under a changing climate in Finland, *Water, Air, and Soil Pollution* **82** (1995), 309–320.

[7] Knoke, T., Stimm, B., Ammer, C., and Moog, M.: Mixed forests reconsidered: A forest economics contribution on an ecological concept, *Forest Ecology and Management* **213** (2005), 102–116.

[8] Knoke, T.: Mixed forests and finance Methodological approaches, *Ecological Economics* **65** (2008), 590–601.

[9] Lohmander, P.: The multi species forest stand, stochastic price process and adaptive selective thinning, *Systems Analysis – Modeling – Simulation* **9** (1992), 229–250.

[10] Robert, C. P., and Casella, G.: *Monte Carlo statistical methods*. Springer, New York, 2004.

[11] Rubinstein, R. Y., and Kroese, D. P.: *Simulation and the Monte Carlo method*. John Wiley & Sons, Hoboken, N.J., 2008.

[12] Sahashi, Y.: Optimal forestry control with variable forest area, *Journal of Economic Dynamics and Control* **26** (2002), 755–796.

[13] Valkonen, S., and Valsta, L.: Productivity and economics of mixed two-storied spruce and birch stands in Southern Finland simulated with empirical models, *Forest Ecology and Management* **140** (2001), 133–149.

Do active labour market policy programmes contribute to lower unemployment in the EU countries?

Jana Hančlová[1], Milan Šimek[2]

Abstract. The paper deals with an estimate of the active labour market policy impact on unemployment rate development for 20 European economies in 2000-2012. The period is split up into the pre-crisis period (2000-2007) and post-crisis period. In a center of research there are 5 types of active labour market policy programmes (training, employment incentives, supported employment and rehabilitation, direct job creation, start-up incentives). For the effect estimate linked to the relative expenditures on those programmes on the unemployment rate development a generalized method of moments has been used where the lagged explanatory variables are applied as the instrumental variables. To investigate the economic crisis effect a dummy variable has been introduced. The analysis results prove that in the pre-crisis period the unemployment decreased mainly as a result of employment incentives but during the crisis an efficiency of this tool fell but training and direct job creation are added.

Keywords: active labour market policy, unemployment rate, economic crisis, panel models, generalized method of moments.

JEL Classification: J68, H53, C54

AMS Classification: 91B40

1 Introduction

Many authors deal with the issue how the European programs supporting active labour market policy are efficient. For example, Kluve [9] points out different effects of particular active labour market policies before 2008. He refers to their various effects in private and public sectors and draws his attention to substantial differences in their implementation in the EU individual member states. Empirical results of the executed meta analysis proved that training programmes have a positive impact on unemployment rate when terminated. Private sector incentive programmes and services and sanctions show a significantly better performance. Evaluations of these types of programmes have a 30 to 50 percentage points higher probability of estimating a significant positive impact than training programmes. On the other hand, evaluation of direct employment programmes is about 25 percentage points less likely to estimate a significant positive impact on post programme employment outcomes.

Sirovátka, Šimíková [13] state that numerous studies from abroad present the fact that labour market policy is just complementary to a complex of macroeconomic interventions regarding a support of recovery as it cannot solve an issue of the lack of the jobs available. European Commission report [5] observes that after the crisis, following the recommendations in the European Economic Recovery Plan of December 2008, reform activity focused on supporting aggregate demand, employment and income support to reduce social distress caused by the crisis whilst at the same time on easing transition to new jobs. For example, Concepción, Congregado and Millán [2] describe the effects of start-up incentives programme during the crisis. Latest analyses of global labour market policy orientation in the EU countries during the crisis [1] identify new alternate orientations of labour market policies which are influenced by interpretation of the issues raised by the impact of the crisis in a way suggested by decisive political actors in the given countries.

Immervoll and Scarpetta in [8] based on their research present that training programmes and even

[1] VSB-Technical University of Ostrava/Faculty of Economics, Department of Systems Engineering, Sokolska 33, Ostrava 1,701 21, Czech Republic, jana.hanclova@vsb.cz.
[2] VSB-Technical University of Ostrava/Faculty of Economics, Department of National Economy, Sokolska 33, Ostrava 1, 70121, Czech Republic, milan.simek@vsb.cz.

publically-subsidised work-experience programmes can help to prevent this group from becoming discouraged and detached, while preparing them to take advantage of new job opportunities once job creation picks up. But indeed a vigorous recovery in employment is key to tackle the current large imbalances in the labour market. Activation and employment support are best seen as "greasing the wheels" of the labour market. Programmes that ease efficient job reallocation improve labour market outcomes irrespective of the economic cycle.

The aim of the paper is to study the effects of various tools of active labour market policy on the unemployment rate in 20 chosen European countries and to compare these results in pre-crisis period 2000-2007 and in the crisis period 2008-2012.

The paper is divided into five sections. The first one specifies the issue of research. The second section defines data and analyses them. The third section deals with a proposal of panel model and a method of its estimate. The fourth section presents the empirical results and related discussion. The final section summarizes the important conclusions of the research.

2 Data analysis

The paper deals with an unemployment rate (*UR*) expressed in % on one hand. On the other hand, there is a percentage share of total expenditures on active labour market policy to gross domestic product (*LMP_C27*). LMP measures covered interventions providing a temporary support for the disadvantaged groups at labour market and aiming at activating the unemployed, encouraging people to move from involuntary inactivity into jobs, or maintaining the jobs for persons threatened by unemployment. The individual tool types of active labour market policy (LMP) which are subject of the research follow: training (*LMP_TR*), employment incentives including job rotation and job sharing (*LMP_EI*), supported employment and rehabilitation (*LMP_SER*), direct job creation (*LMP_DJC*), start-up incentives (*LMP_SUI*).

The given features of active labour market policy are monitored in 20 European countries: Belgium (BE), Bulgaria (BG), Czech Republic (CZ), Denmark (DK), Germany (DE), Ireland (IE), Greece (EL), Spain (ES), France (FR), Latvia (LV), Lithuania (LT), Luxembourg (LU), Austria (AT), Poland (PL), Portugal (PT), Slovenia (SI), Slovakia (SK), Finland (FI), Sweden (SE), United Kingdom (UK). Another European countries were not included due to the lack of empirical data on unemployment rate or LMP expenditures as per an individual tool type. All data used come from the Eurostat database [6] over 2000-2012.

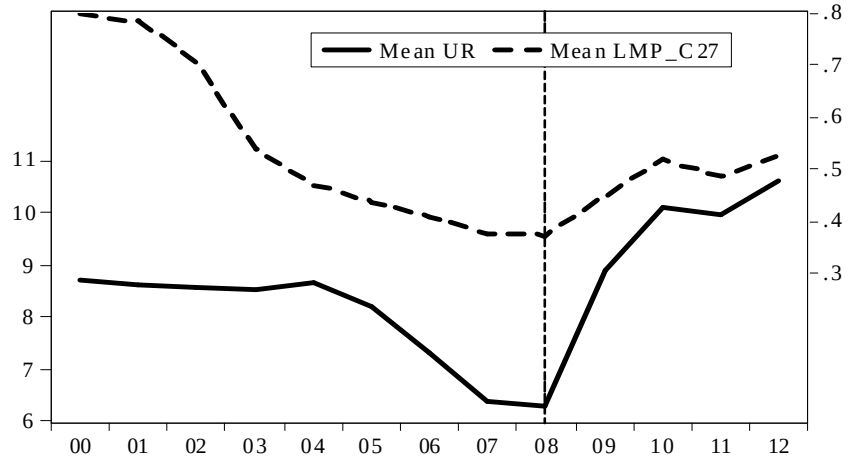


Figure 1 Development of unemployment rate and relative expenditures on active LMP

Figure 1 shows the development of unemployment rate (mean UR) and development of percentage share of expenditures on active labour market policy to GDP (mean LMP_C27). The values are set across the economies in the centre of research in 2000-2012. In the pre-crisis period 2000-2007 mean LMP_C27 slightly decreased from 0.80% in 2000 to 0.39% in 2007 (a scale on the right axis). During crisis this share raised progressively to 0.53% in 2012. If we examine the expenditures on these tools of active labour market policy in EUR we can declare that a volume of spent funds was more less the same until 2008 and a decrease of the share of expenditures happened due to the GDP growth in the EU till 2007. During crisis, mainly in 2009 and 2010, a significant increase of absolute expenditures on active labour market policy was recorded. A similar development was registered for an indicator mean rate of unemployment

(mean UR). In period 2000-2007 it fell from 8.71% in 2000 to 6.29% in 2008. Between 2008-2010 it sharply raised to 10.11% and following a moderate stagnation it went up to 11.42% in 2012. Tvrdon and Verner [15] look into a detailed process of regional unemployment disparities and their dynamics, too. The authors Szomolányi, Lukáčik and Lukáčiková [14] pay attention to the labour market influence on economy performance in Visegrad Four Countries.

Regarding a proportion of expenditures on individual tools of active labour market policy in the monitored period the expenditures on training and retraining programmes predominated. Employment incentives and supported employment and rehabilitation were also represented significantly. Figure 2 shows development of relative expenditures on individual tool types in the monitored period.

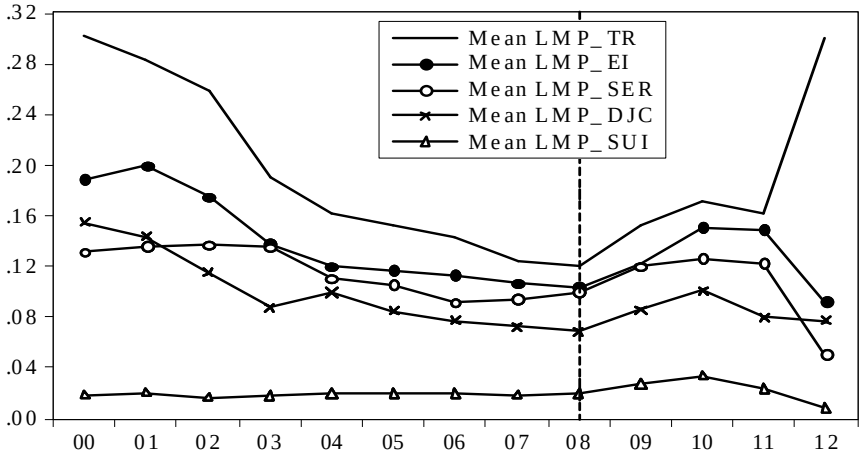


Figure 2 Development of relative expenditures on individual tool types of active LMP

A graphical data analysis shows that in a long-term there are significant differences in the active labour market financing in individual European economies. These differences are connected with the economic policy priorities [7] and an extent to which the domestic economy was hit during crisis. Higher expenditure shares from GDP on active labour market policy are in a long-term registered in Denmark, Belgium, Sweden, Finland and France. Very low share from GDP on active labour market policy is alocated in Bulgaria, Romania, Lithunia, Latvia, Czech Republic, Slovakia and Slovenia.

3 Formulation and a panel generalized method of moments

When modelling the impact of active labour market policy tools on unemployment development we formulate a panel model monitoring also the differences of a tool efficiency in pre-crisis and post-crisis periods; it means splitting of the period into two groups, i.e 2000-2007 and 2008-2012. We introduce dummy variables in a multiplicative form. Formulation of the panel model is given by the equation:

$$\begin{aligned} UR_{it} &= c_{it} + \sum_{j=1}^{N} \alpha_j LMP\_x_{it}^{j} + \sum_{j=1}^{N} \beta_j LMP\_x_{it}^{j}\_dc + \varepsilon_{it} = c_{it} + \\ &+ \alpha_{TR} LMP\_TR_{it} + \alpha_{EI} LMP\_EI_{it} + \alpha_{SER} LMP\_SER_{it} + \alpha_{DJC} LMP\_DJC_{it} + \alpha_{SUI} LMP\_SUI_{it} + \\ &+ \beta_{TR} LMP\_TR_{it}\_dc + \beta_{EI} LMP\_EI_{it}\_dc + \beta_{SER} LMP\_SER_{it}\_dc + \beta_{DJC} LMP\_DJC_{it}\_dc + \\ &+ \beta_{SUI} LMP\_SUI_{it}\_dc + \varepsilon_{it}, \end{aligned} \tag{1}$$

where index $j = 1,...,N$ represents tool type of active labour market policy, *dc* is dummy variable for a division into pre-crisis period 2000-2007 (*dc=0*) and crisis period 2008-2012 (*dc=1*). The slope change is reflected by parameter β_j for variables:

$$LMP\_x_{it}^{j}\_dc = LMP\_x_{it}^{j} \cdot dc_{it}. \tag{2}$$

A Panel Generalized Method of Moments (PGMM) was chosen for the model estimate where one period lagging explanatory variables $LMP\_x_{it-1}^{j}$ and $LMP\_x_{it-1}^{j}\_dc$ were used as instrumental variables. When specifying an estimator of the Generalized Method of Moments (GMM) pay attention to issues of weighting matrix estimation and coefficient covariance calculation. There is restriction of moment conditions that may be written as an

orthogonality condition between the residuals and a set of instruments. As with other instrumental variable estimators, for the GMM estimator to be identified, there must be at least as many instruments as there are parameters in the model. In models where there are the same number of instruments as parameters, the value of the optimized objective function is zero. If there are more instruments than parameters, the value of the optimized objective function will be greater than zero. The value of the objective function, termed the *J* –statistic, can be used as a test of over-identifying moment conditions. The second important aspect of specifying a GMM estimator is the choice of the weighting matrix. Software EViews offers the White weighting matrix which is a heteroskedasticity consistent estimator of the long-run covariance matrix based on an initial estimate of parameters. Nevima and Melecky [11] also devoted their attention to panel data modelling for Visegrad Four Countries.

4 Empirical results and discussion

When estimating the model in equation (1), first of all, we dealt with the **lag** of explanatory variables and then, with the **estimate** of the best model both from the econometric and economic point of view and we discussed the final results.

Lag setting

With regard to the annual data we can introduce explanatory variables in the model (1) in the same time period resp. monitor their effects with a one-year lag. Table 1 sums up paired correlations for variables the unemployment and individual tools of LMP in the same time (upper part of the table) and with a one-year lag of explanatory variables. Further on, rows draw a line among three options – pre-crisis period, post-crisis period and whole period 2000-2012. Statistically significant correlation coefficients at 10% resp. 5% level of significance are indicated by * resp. **. The gained results prove that it is appropriate to choose explanatory variables in the same time as unemployment rate. A paired correlation is statistically significant and negative mainly for LMP tool called supported employment and rehabilitation (SER) and thus, in pre-crisis period it influences more intensively an unemployment decrease. Tools as employment incentives (EI) and direct job creation (DJC) are not significantly linearly correlated with unemployment rate.

UR/	***LMP_TR***	***LMP_EI***	***LMP_SER***	***LMP_DJC***	***LMP_SUI***
2000 - 2007	-0.181*	-0.087	-0.307**	+0.004	+0.364**
2008 - 2012	-0.008	-0.065	-0.295*	+0.114	+0.390**
2000 - 2012	-0.117*	-0.081	-0.290**	+0.027	+0.390**
UR/	***LMP_TR(-1)***	***LMP_EI(-1)***	***LMP_SER(-1)***	***LMP_DJC(-1)***	***LMP_SUI(-1)***
2000 - 2007	-0.143	-0.052	-0318**	+0.045	+0.389**
2008 - 2012	-0.038	-0.066	-0.287**	+0.087	+0.456**
2000 - 2012	-0.115	-0.076	-0.284**	-0.025	+0.435**

Table 1 Correlation analysis of UR and active LMP tools

Model estimate and empirical result discussion

The results of the panel model estimate in equation (1) using the Generalized Method of Moments are summed up in Table 2 for 12 periods and 20 European economies where all necessary data were available. The model includes just variables with a statistically significant regression parameter at 5% level of significance. Table 2 includes the standard coefficient estimates, standard errors, t-statistics and associated *p*-values. Concerning the effect of individual active labour market policy tools we can prove that:

- **Training *(TR)*** causes an increase of unemployment in pre-crisis period 2000-2007 – with an increase of expenditure share on training/GDP by 1% there is a temporary unemployment growth in the same year by 7.392%, ceteris paribus. But in crisis period 2008-2012 this tool became already efficient. With an increase of expenditure share on training/GDP by 1% an unemployment decrease began to realize by 2.296% (i.e. 7.392-9.588=-2.296%). These conclusions are in accordance with the research of Kluve [10], where an author dealt with meta-analysis of efficiency as concerns 137 active labour market programmes in 19 European countries. Conclusions demonstrate that training programmes - as the most commonly used type of active policy - show modestly positive effects mainly till eighties and later there is a minor or negative impact in Sweden, Norway or the United Kingdom for juniors.
- **Employment incentives *(EI)*** are the most efficient tool in pre-crisis period - with an increase of expenditure share

on EI/GDP by 1% the unemployment rate is cut down by 35.609%. On the other hand, in comparison with the training tool the efficiency is reduced in crisis but employment incentives still have some effect on lowering of unemployment. In crisis period with an increase of expenditure share on EI/GDP by 1% the unemployment is reduced just by 6.784 % (-35.609+28.825=-6.784). Authors Kluve, Lehman and Schmidt [9] studied the effect of a traning programme and a wage subsidy scheme for the Polish Labor Force Survey in 1990-1996. Their findings suggest that training raises *individual* employment probability, while wage subsidies display negative treatment effects for participants in the Polish case. These conclusions are in compliance with an agregate employment for training programme in crisis period. Nevertheless, the results concerning employment incentives in our research confirmed a lowering of agregate unemployment. Likewise Eriksson and Pytlikova [4] paid attention to the minimum-wage increases having a positive impact on employment level increase in the Czech Republic and Slovak Republic at the end of nineties.

- **Direct job creation (DJC)** as the active labour market policy tool has a major effect in crisis mostly, but in pre-crisis period DJC is not statistically efficient. It was proved that with an increase of expenditure share on DJC/GDP by 1% the unemployment rate decreases by 15.598% in crisis.
- Active labour market policy tools **supported employment and rehabilitation** (SER) and **start-up incentives** (SUI) do not statistically significantly influence a development of unemployment rate in the same year despite the fact that from the point of paired correlation only they ranked to the statistically significant. The paper of Dana and Emőke-Szidónia [3] deals in detail with the innovation effect in doing business on labour market dynamics and lagging in the Central and Eastern economies in 2007-2011. The results show that innovations in businesses do not generate positive effects in the labour market simulaneously but just with a delay of minimum three years. These results justify why start-up incentives tool does not appear to be efficient. In next research it will be suitable to examine longer delays. In addition, the article of authors Román, Congregado and Millán [2] is devoted to new start-up programmes of active labour market policy in recession (1994-2001). This study suggests that the contribution of these incentives is dubious, if aimed to combat economic and jobs crises as part of the entrepreneurship policy, and can be shaped by various country-specific factors, such as the economic situation and the stringency of labour laws. Regarding further time line we can conclude that start-up incentives tool plays greater role in labour market than at the end of nineties. The study of Parello [12] documents that start-up incentives raise probability of self-employment in recession if there are no unexpected institutional threats. However, this tool is applicable especially when unemployment is moderate.

Cross-sections included: 20		Periods included: 12		
Total panel (unbalanced) observations: 130				
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
C	11.152	0.859	12.989	0.000
LMP_TR	7.392	3.466	2.133	0.036
LMP_EI	-35.609	10.763	-3.308	0.001
LMP_TR_DC	-9.588	3.204	-2.992	0.004
LMP_EI_DC	28.825	6.479	4.449	0.000
LMP_DJC_DC	-15.598	4.692	-3.324	0.001

Table 2 The final model estimate (1)

Coefficient of determination R-squared was at relatively acceptable level 0.78. Statistically significant paired correlation between explanatory variables was identified only for *LMP_EI* and *LMP_DJC* 0.51** and further on, for a couple *LMP_EI* and *LMP_SER* 0.34**. It confirms ortogonality of explanatory variables. The results of the Sargan *J*-statistic 5.88 (value of the GMM objective function at estimated parameters) and the instrument rank 41 is greater than the number of estimated coefficients (5), we may use it to construct the Sargan test of over-identifying restrictions. It is worth noting that the *J*-statistic reported by a panel equation differs from that reported by an ordinary equation by a factor equal to the number of observations. Under the null hypothesis that the over-identifying restrictions are valid. The *p*-value of *J*-statistic is 0.32 and we reject the over-identifying restrictions.

5 Conclusions

The objective of the paper was to examine the effect of active labour market policy tools on the unemployment rate development in chosen European countries in pre-crisis period (2000-2007) and post-crisis period (2008-2012).

The results document that the effect of active labour market policy tools is mostly evident in the same year. The

panel model estimate for 20 European countries demostrated that:

- Employment incentives are the most efficient active labour market policy tool in pre-crisis period but in crisis the efficiency decreases but they still lower rate of unemployment.
- Subsequent tool in discussion is training which more likely supports unemployment in pre-crisis period but in crisis it is the third most efficient tool to reduce unemployment.
- Direct job creation is another important tool which is not much evident in pre-crisis period but in crisis it is the second most efficient tool to decrease unemployment.
- Additional tools as supported employment and rehabilitation and start-up incentives do not confirm major effect on cut of unemployment over whole period of research.

Results present that in pre-crisis period the tools orientated on labour market supply side have some effect. They become more efficient in crisis and in addition, tools focused on demand side are also applied. This research has not taken into consideration age categories of unemployed or their educational structure. Thus, it was not possible to study specific features of juniors or seniors.

Acknowledgements

This research was supported by the Student Grant Competition of Faculty of Economics, VSB-Technical University of Ostrava within projects SP2014/131 and SP2014/113 and also by the European Social Fund within the project CZ.1.07/2.3.00/20.0296.

References

[1] Clasen J., Clegg, D., and Kvist, J. European labour market policies in (the) crisis. *Working paper 12/2012*. Brussels: ETUI, 2012.
[2] Concepcion, R., Congregado E., and Millán, J. M.: Start-up incentives: Entrepreneurship policy or active labour market programme? *Journal of Business Venturing* **28** (2013), 151-175.
[3] Dana, N.A.R., and Emőke-Szidónia, F.: Labour market dynamics as time-lagged effect of entrepreneurship in the case of Central and Eastern European countries. *Procedia Economics and Finance* **3** (2012), 950-955. Directorate-General for Economic and Financial Affairs, 2012.
[4] Eriksson, T., and Pytlikova, M.: Minimum-wage increases and employment: empirical evidence from the Czech and Slovak Republics. *Finance a úvěr – Czech Journal of Economics and Finance* **52** (2002), 614-615.
[5] European Commission: *Labour market developments in Europe 2011*. Brussels: European Commission, 2011.
[6] Eurostat: Eurostat database. [online], available at http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database *[cit. 2014-03-25]*
[7] Hanclova, J., Simek, M., and Horak, J.: Factors influencing the long-term unemployment level and development in the European Union. In: *Proceedings of 30th International Conference on Mathematical Methods in Economics*. Silesian University, Karviná, 2012, 249-254.
[8] Immervoll, H., and Scarpetta, S.: Activation and employment support policies in OECD countries. An overview of current approaches. *IZA Journal of Labor Policy* **19** (2012), 1-20.
[9] Kluve, J., Lehmann, H., and Schmidt Ch. M.: Disentangling treatment effects of active labor market policies: the role of labor force status sequences. *Labour Economics* **15** (2008), 1270-1295.
[10] Kluve, J.: The effectiveness of European active labor market programs. *Labour Economics* **17** (2010), 904-918.
[11] Nevima, J., and Melecky, L.: Regional Competitiveness Evaluation of Visegrad Four Countries through Econometric Panel Data Model. In: *Proceeding of 10th International Conference of Liberec Economic Forum*. Technical University of Liberec, Liberec, 2011, 348-361.
[12] Parello C. P.: Labor market rigidity and productivity growth in a model of innovation-driven growth. *Economic Modelling* **28** (2011), 1058-1067.
[13] Sirovátka, T. and Šimíková, I. *Politika zaměstnanosti a další opatření na trhu práce v dlouhodobé perspektivě a v průběhu krize.* Praha: Výzkumný ústav práce a sociálních věcí, 2013.
[14] Szomolányi, K., Lukáčik, M. and Lukáčiková, A.: Estimation of the Production Function Parameters in V4 Economies. In: *Proceedings of 31st International Conference on Mathematical Methods in Economics*. College of Polytechnics, Jihlava, 2013, 898-902.
[15] Tvrdon, M., and Verner, T.: Regional unemployment disparities and their dynamics: Evidence from the Czech Republic. In: *Proceedings of 30th International Conference on Mathematical Methods in Economics*. Silesian University, Karviná, 2012, 938-943.

Iterative method for bandwidth selection in kernel discriminant analysis

Kamila Hasilová[1]

Abstract. Kernel estimates belong to very effective nonparametric estimates of a probability density function. This concept provides an interesting alternative to the classical parametric approach in empirical studies. Kernel estimates depend on a bandwidth matrix that controls smoothness of the estimated density.

This paper is focused on the use of kernel estimates with the bandwidth matrix computed by the so-called 'iterative method' in kernel discriminant analysis. Its utility is illustrated through a short simulation study and real data applications, where standard model-based discrimination rules are compared with a kernel discriminant rule.

Keywords: Kernel density estimate, iterative method, discriminant analysis.

JEL classification: C14
AMS classification: 62G07, 62H30

1 Discriminant analysis

The basic problem of discrimination is to allocate a new observation $\mathbf{z}$ into one of m distinct populations or groups, denoted as $G_1, \ldots, G_m$, with prior probabilities $\pi_1, \ldots, \pi_m$ and corresponding density functions $f_1, \ldots, f_m$ [9]. A standard approach to this problem is using Bayes discriminant rule

$$\mathbf{z} \text{ is assigned to group } G_k \text{ if } k = \underset{i=1,\ldots,m}{\arg\max}\ \pi_i f_i(\mathbf{z}).$$

The unknown prior probabilities and the unknown density functions are estimated from training data $\{\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i}\}$ with sample sizes n_i (known and non-random), coming from f_i for $i = 1, \ldots, m$. Using this discriminant rule, we allocate the test data $\mathbf{Z}_1, \ldots, \mathbf{Z}_\ell$ drawn from density $f = \sum_{i=1}^m \pi_i f_i$.

The classical approach to discriminant analysis substitutes a particular parametric form for f_i into the discriminant rule. The well-known parametric approaches are linear and quadratic discriminant analysis based on the multivariate normality assumption [12]. If we assume a common covariance matrix $\mathbf{\Sigma}$ and possibly different mean vectors $\boldsymbol{\mu}_i$, it gives the linear discriminant rule (after taking logarithms of f_i)

$$\text{LDR:}\quad \mathbf{z} \text{ is assigned to group } G_k \text{ if } k = \underset{i=1,\ldots,m}{\arg\max}\ \ln\pi_i - \tfrac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_i)^T\mathbf{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu}_i).$$

The common covariance matrix is estimated using a pooled estimate of the covariance $\widehat{\mathbf{\Sigma}} = \big(\sum_{i=1}^m n_i - m\big)^{-1}\big[(n_1-1)\widehat{\mathbf{\Sigma}}_1 + \cdots + (n_m-1)\widehat{\mathbf{\Sigma}}_m\big]$, where $\widehat{\mathbf{\Sigma}}_i$ is the sample covariance matrix from the training sample for the group G_i. Allowing different covariance matrices yields the quadratic discriminant rule

$$\text{QDR:}\quad \mathbf{z} \text{ is assigned to group } G_k \text{ if } k = \underset{i=1,\ldots,m}{\arg\max}\ \ln\pi_i - \tfrac{1}{2}\ln|\mathbf{\Sigma}_i| - \tfrac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_i)^T\mathbf{\Sigma}_i^{-1}(\mathbf{z}-\boldsymbol{\mu}_i).$$

Generalisation of these parametric methods to the nonparametric one is straightforward. Instead of assuming a normal distribution of the underlying densities, we use kernel estimates of these densities based on the training data [11]. The kernel discriminant rule is

$$\text{KDR:}\quad \mathbf{z} \text{ is assigned to group } G_k \text{ if } k = \underset{i=1,\ldots,m}{\arg\max}\ \pi_i \hat{f}_i(\mathbf{z}, \mathbf{H}_i), \tag{1}$$

where $\hat{f}_i(\mathbf{z}, \mathbf{H}_i)$ is the kernel density estimate corresponding to the group G_i.

[1]University of Defence, Faculty of Economics and Management, Department of Econometrics, Kounicova 65, Brno 662 10, Czech Republic, kamila.hasilova@unob.cz

2 Kernel density estimation

Let a d-variate random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ come from a continuous probability distribution with a density f. The kernel density estimator $\hat{f}$ is defined as a weighted average of observations at a point $\mathbf{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$

$$\hat{f}(\mathbf{x}, \mathbf{H}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i) = \frac{1}{n} |\mathbf{H}|^{-1/2} \sum_{i=1}^{n} K\big(\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{X}_i)\big).$$

K is a d-variate kernel function, which is often taken to be a probability density function satisfying $\int_{\mathbb{R}^d} K(\mathbf{x})\, \mathrm{d}\mathbf{x} = 1$ (we omit the subscript $\mathbb{R}^d$ in the rest of the text). $\mathbf{H}$ is a $d \times d$ symmetric positive definite matrix called a bandwidth matrix with $|\mathbf{H}|$ standing for its determinant.

The choice of the kernel is not crucial, so the standard normal kernel, i.e., $K(\mathbf{x}) = (2\pi)^{-\frac{d}{2}} \exp(-\frac{1}{2}\mathbf{x}^T\mathbf{x})$, is the most commonly used kernel. In contrast, selection of the bandwidth parameter plays a key role in kernel density estimation due to its effect to accuracy of the estimate. Although there are more smoothing parameters to select and the sparseness of the data in the higher-dimensional space makes the estimation process more difficult than in the univariate case, several studies have confirmed that the kernel density estimator is an effective tool for visualization of two- and three-dimensional data sets [10, 14].

The most common measure of quality of the estimate $\hat{f}$ is the Integrated Square Error $\mathrm{ISE}(\mathbf{H}) = \int [\hat{f}(\mathbf{x}, \mathbf{H}) - f(\mathbf{x})]^2\, \mathrm{d}\mathbf{x}$, which gives the distance between the estimate $\hat{f}$ and the underlying density f. As this measure is a random variable, it is useful to take into account its expectation $\mathrm{MISE}(\mathbf{H}) = E\,\mathrm{ISE}(\mathbf{H})$. Even if the standard bias-var decomposition $\mathrm{MISE}(\mathbf{H}) = \int \mathrm{var}\, \hat{f}(\mathbf{x}, \mathbf{H})\, \mathrm{d}\mathbf{x} + \int \mathrm{bias}^2\, \hat{f}(\mathbf{x}, \mathbf{H})\, \mathrm{d}\mathbf{x}$ is used, it is easily seen that finding the bandwidth matrix $\mathbf{H}_{\mathrm{MISE}}$, which minimizes this error, is very difficult. Wand and Jones [14] derived (under some assumptions on the density f, kernel function K and bandwidth matrix $\mathbf{H}$) the Asymptotic Mean Integrated Square Error (AMISE)

$$\mathrm{AMISE}(\mathbf{H}) = \underbrace{n^{-1}|\mathbf{H}|^{-1/2} V(K)}_{\mathrm{AIV}} + \underbrace{\tfrac{1}{4} m_2^2(K) (\mathrm{vec}\, \mathbf{H})^T V(\mathrm{vec}\, D^2 f)\, \mathrm{vec}\, \mathbf{H}}_{\mathrm{AISB}},$$

where the following notation is used [2, 13]

- $V(g) = \int g(\mathbf{x}) g^T(\mathbf{x})\, \mathrm{d}\mathbf{x}$ for any square integrable vector valued function g;
- $\int \mathbf{x}\mathbf{x}^T K(\mathbf{x})\, \mathrm{d}\mathbf{x} = m_2^2(K) \cdot \mathbf{I}_d$;
- vec is a vector operator that transforms matrix into a vector by stacking all the columns of the matrix underneath each other in order from left to right;
- D^2 is a standard Hessian operator, i.e., a matrix operator of the second partial derivatives;
- AIV stands for an asymptotic integrated variance and AISB stands for an asymptotic integrated square bias.

AMISE provides a basis for an asymptotic analysis. However, it is not possible, in general, to compute the bandwidth matrix $\mathbf{H}_{\mathrm{AMISE}}$, which minimizes the AMISE error, because the AISB part depends on the underlying density f. It has led to a search for estimating AMISE and also MISE.

The plug-in methods are based on substitution of the bias term in AMISE. Elements of $V(\mathrm{vec}\, D^2 f)$ are estimated using a pilot bandwidth matrix (different from the matrix $\mathbf{H}$). It produces an estimate of AMISE which can be numerically minimized [3]. The least square cross-validation method is targeting MISE and uses the so called 'leave-one-out' estimator of density function. Although this method provides an unbiased estimate of MISE, it suffers from large variability in results, which might be expected based on experience in the univariate setting [14]. The simplest way to select the bandwidth matrix is given by a normal reference rule. It is presumed that the underlying density f is the d-variate Gaussian density [2, 10]. Such an assumption might be restrictive, on the other hand, the normal reference rule provides a useful initial estimate in the numerical minimization.

The iterative method, proposed by Horová and her team ([4, 5]), is based on finding a solution of the equation $\mathrm{d}\,\mathrm{AMISE}(\mathbf{H}) \big/ \mathrm{d}\,\mathrm{vec}\, \mathbf{H} = 0$, which is equivalent to a minimization of AMISE with respect to $\mathbf{H}$. It can be shown that the AMISE-optimal bandwidth matrix $\mathbf{H}_{\mathrm{AMISE}}$ fulfills the following relationship between the asymptotic integrated variance (AIV) and the asymptotic integrated square bias (AISB)

$$d \cdot \mathrm{AIV}(\mathbf{H}) = 4 \cdot \mathrm{AISB}(\mathbf{H}).$$

A	G_1: $N_2(0,0;1,1,0)$	G_2: $N_2(1,1;1,1,0)$
B	G_1: $N_2(1/2,1/2;5/4,1/4,0)$	G_2: $N_2(0,0;3/2,1/4,-9/10)$
C	G_1: $1/3N_2(-1,-1;1,1,0) + 1/3N_2(1,-2;1,1,0)$ $+1/3N_2(4,0;1,1,0)$	G_2: $N_2(1,1;1,1,0)$
D	G_1: $Wb(1,3)\cdot Wb(1,2)$	G_2: $Wb(3/2,2)\cdot Wb(2,3)$

Table 1 Parameters of the simulated situations ($N_2(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2,\rho)$ and $Wb(\lambda,k)$)

Horová *et al.* [4] used estimates of AIV and AISB of the form

$$\widehat{\mathrm{AIV}}(\mathbf{H}) = \frac{1}{n}|\mathbf{H}|^{-1/2}V(K),$$
$$\widehat{\mathrm{AISB}}(\mathbf{H}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(K_{\mathbf{H}} * K_{\mathbf{H}} * K_{\mathbf{H}} * K_{\mathbf{H}} - 2K_{\mathbf{H}} * K_{\mathbf{H}} * K_{\mathbf{H}} + K_{\mathbf{H}} * K_{\mathbf{H}})(\mathbf{X}_i - \mathbf{X}_j),$$

where $*$ denotes a convolution operation. It gives a basic equation of the iterative method

$$dn|\mathbf{H}|^{-1/2}V(K) = 4\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(K_{\mathbf{H}}^{*4} - 2K_{\mathbf{H}}^{*3} + K_{\mathbf{H}}^{*2})(\mathbf{X}_i - \mathbf{X}_j).$$

This is an equation of a variety in $\mathbb{R}_+^d$, so entries h_{ij} of the bandwidth matrix $\mathbf{H}$ are not uniquely determined. According to the normal reference rule, the entries of $\mathbf{H}$ can be expressed as suitable multiples of the sample covariance matrix. Also it can be expected that it is possible to express every entry of the bandwidth matrix as a multiple of its first entry h_{11}, i.e., $h_{ij} = c\cdot h_{11}$ for $i,j = 1,\ldots,d$. Combining these ideas they came with supplementary equations for the entries of $\mathbf{H}$ [13]

$$h_{ij} = \frac{s_{ij}}{s_{11}}h_{11} \quad \text{for} \quad i,j = 1,\ldots,d,$$

where s_{ij} denotes the elements of the sample covariance matrix $\mathbf{S}$. The iterative method for the bandwidth matrix becomes

$$nd|\mathbf{H}|^{-1/2}V(K) = 4\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(K_{\mathbf{H}}^{*4} - 2K_{\mathbf{H}}^{*3} + K_{\mathbf{H}}^{*2})(\mathbf{X}_i - \mathbf{X}_j) \quad \text{and} \quad h_{ij} = \frac{s_{ij}}{s_{11}}h_{11} \quad (i,j = 1,\ldots,d).$$

This nonlinear equation for the unknown h_{11} can be solved by an appropriate iterative numerical method, which gives the name to the method.

3 Simulated and real data examples

In this section, we compare the linear and quadratic discriminant rules with the discriminant rule based on the kernel density estimates.

We have considered the following situations (see Table 1 and Figure 1), where groups are of the size 100 each. Situations A–C consist of two groups coming from a normal distribution. Both groups of A have the same covariance matrix. In B, groups differ in their covariance matrices; the group G_1 of C is a mixture of three normal densities. Situation D is formed by two groups with the Weibull distribution.

We also need to measure the quality of the classification rules. To test the accuracy of the procedure one can use a proportional error statistics

$$\tau = \frac{n_c - \sum_{i=1}^{m} p_i n_i}{n - \sum_{i=1}^{m} p_i n_i},$$

where n_c is the number of cases correctly classified, p_i is the estimated prior probability π_i, n_i is the number of cases in group G_i $(i = 1,\ldots,m)$, and n is the total number of cases over all the groups. This

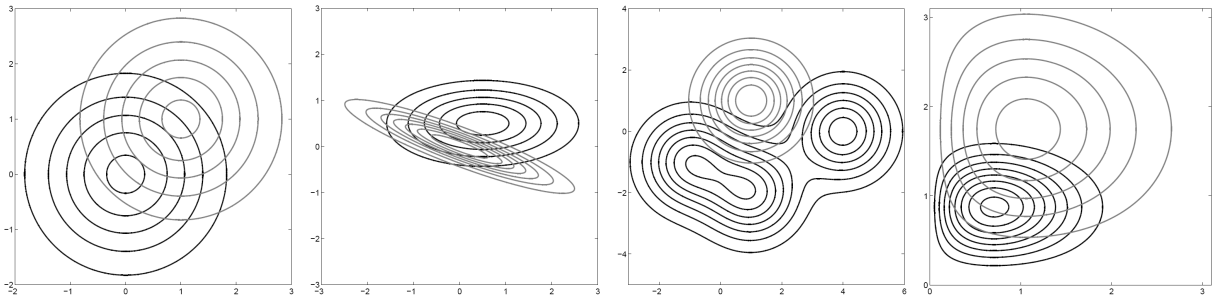


Figure 1 Contour plots of simulated densities for situations A–D

Situation		A	B	C	D
KDR_{IT}	mean (std)	0.5149 (0.0547)	0.6727 (0.0467)	0.7983 (0.0505)	0.6411 (0.0558)
LDR	mean (std)	0.5319 (0.0517)	0.6273 (0.0569)	0.6400 (0.0483)	0.6214 (0.0513)
QDR	mean (std)	0.5309 (0.0537)	0.6791 (0.0410)	0.7892 (0.0442)	0.6506 (0.0518)

Table 2 Means and standard deviations of τ coefficient for simulated situations

coefficient gives us a standardized measure of improvement with respect to a pure random assignment. Values of τ range from 1, there are no errors in classification, to 0, which indicates no improvement. It is also possible to get negative results, which indicates no discrimination. Because n_c must be an integer, the numerator could become slightly negative—situation with no group differences [8].

Application of the classification rule to the same data from which it has been formed tends to overestimate the power of the classification procedure [9]. To avoid too optimistic assessment of the error rate, we have randomly split the sample into two subsets of the same size and used one subset to derive the classification functions and the other one to test the classifications. We have repeated this 100 times for each situation. The means and standard deviations of τ for kernel discriminant rule with bandwidth matrices selected according to the iterative method (KDR_{IT}) and for linear and quadratic discriminant rules (LDR, QDR) are summarized in Table 2.

We may conclude that there is no improvement in the situation A in using the kernel discriminant rule instead of the classical linear/quadratic discrimination. Similar behaviour can be seen for the situation B. The overall discrimination is better than in the situation A and, as it can be expected, the quadratic discriminant rule gives better results than the linear one. The situation C, where one population is partially encircled by the other one, shows a light advance in using kernel discrimination rule with bandwidth matrix selected by the iterative method over the parametric discrimination rules. In the case of Weibull densities, the situation D, we can see that the nonparametric discrimination rule is comparable to the linear and quadratic discriminant rules.

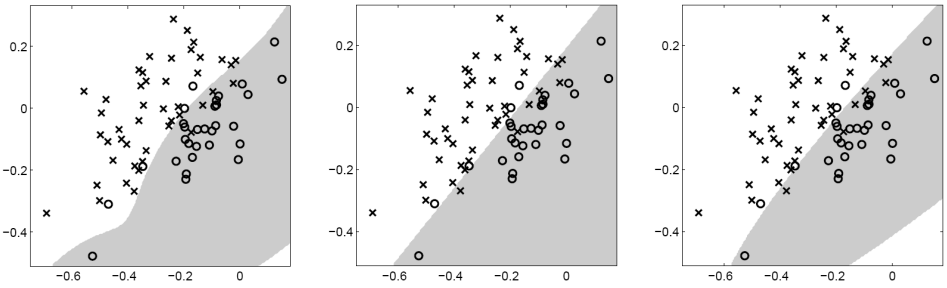
Although it seems that there is little improvement in using kernel discriminant rules, the real data sets give a different view on discrimination and its results.

We use several data sets to illustrate the kernel discriminant rule. Let us start with the haemophilia data set, which contains two measured variables on 75 women belonging to two groups. Group A of the size $n_A = 30$ are non-carriers (normal group) and group B of the size $n_B = 45$ are known haemophilia A carriers (obligatory carriers). Discrimination is based on two feature variables: $\log_{10}$(AHF activity) and $\log_{10}$(AHF antigen) (see, e.g., [6]).

However, τ coefficient cannot be used in this situation because we do not have training and test groups. Instead of τ we employ the misclassification rate [3]

$$\text{MR} = 1 - \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n_j} I[\mathbf{X}_{ij} \text{ is correctly classified using } \text{KDR}_{-ij}],$$

where KDR_{-ij} is the kernel discrimination rule in equation (1) except that π_i and $\hat{f}_i(\mathbf{z}, \mathbf{H}_i)$ are replaced by their cross-validation (jack-knife) counterparts. In other words, we omit one observation, develop a



(a) $\mathrm{KDR_{IT}}$ (b) LDR (c) QDR

Figure 2 Classification regions for haemophilia data set: ∘ grey – non-carriers, × white – carriers

Data	haemophilia	CD rate	bank notes	bankruptcy	Iris
$[d; m]$	$[2; 2]$	$[1; 2]$	$[2; 2]$	$[4; 2]$	$[4; 3]$
$\mathrm{MR(KDR_{IT})}$	0.1333	0.3043	0.0050	0.1087	0.0333
MR(LDR)	0.8267	0.2754	0.0150	0.5652	0.0200
MR(QDR)	0.4533	0.2899	0.0100	0.4130	0.0267

Table 3 Misclassification rates for real data sets (d = dimension of the data, m = number of groups)

discrimination rule based on $n - 1$ observations, and classify the holdout observation [7]. From Table 3 we can conclude that the kernel discrimination rule gives the best result for the haemophilia data. Considering the parametric discrimination rules, the quadratic rule is better than the linear one. This can be seen also from Figure 2.

The kernel discriminant analysis is suitable not only for two-dimensional data, it can be applied to one-dimensional as well as to higher dimensional data sets. The three-month certificate of deposit (CD) rates represent one-dimensional data. The data consist of 69 returns on CD for Long Island banks (group A, $n_A = 29$) and thrifts (group B, $n_B = 40$) [12]. The misclassification rates in Table 3 show that LDR is the best discrimination rule for this data which could be expected since means and standard deviation of both groups are close to each other: $\overline{x}_A = 8.15$, $s_A = 0.32$ and $\overline{x}_B = 8.35$, $s_B = 0.25$.

The next data set is taken from [12], where it is also analyzed. The data comprises the characteristics of 100 real and 100 forged Swiss bank notes. Discrimination is based on the bottom margin width and image diagonal length of the bills (in millimetres). The classification is comparable for all three methods and overall classification is quite good (see Table 3).

In [7], there are provided annual financial data that had been collected for bankrupt firms (group A, $n_A = 21$) approximately two years prior to their bankruptcy and for financially sound firms (group B, $n_B = 19$) at about the same time. The discrimination is based on four variables: x_1 = cash flow / total debt, x_2 = net income / total assets, x_3 = current assets / current liabilities, and x_4 = current assets / net sales. One can observe similar pattern for misclassification rate as for the haemophilia data – Table 3 – KDR gives the best result among studied discrimination rules.

The last data set, which we take a look at, is the well-known Fisher's Iris data set [1]. Although we have compared the discrimination rules only on two-group classification, it can be easily modified to more groups. The Iris data set consists of three groups of 50 observations each and the discriminant variables are widths and lengths of petals and sepals measured in centimetres. From the last column of Table 3 we can see that the overall classification is quite good and all studied methods provide similar results.

4 Conclusion

We have presented the iterative method for computing the bandwidth matrix for kernel density estimates. We have illustrated its use in kernel discriminant analysis, where we have compared nonparametric kernel discriminant rule with the standard parametric rules. The simulations and real data application have

showed that the kernel discriminant rule is more flexible than the usual linear and quadratic rules.

Acknowledgements

This research has been supported by the University of Defence through the Institutional development project Economic Laboratory.

References

[1] Bache, K., Lichman, M.: *UCI Machine learning repository* `http://archive.ics.uci.edu/ml`. Irvine, 2013.

[2] Chacón, J. E., Duong, T., Wand, M. P.: Asymptotics for general multivariate kernel density derivative estimators, *Statistica Sinica* **21** (2011), 807–840.

[3] Duong, T.: ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R, *Journal of Statistical Software* **21**, 7 (2007), 1–16.

[4] Horová, I., Koláček, J., Zelinka, J., Vopatová, K.: Bandwidth choice for kernel density estimates. In: *Proceedings of IASC* 2008, 542–551.

[5] Horová, I., Zelinka, J.: Contribution to the bandwidth choice for kernel density estimates, *Computational Statistics* **22** (2007), 31–47.

[6] Hubert, M., van Driessen, K.: Fast and robust discriminant analysis, *Computational Statistics* **45** (2004), 301–320.

[7] Johnson, R. A., Wichern, D. W.: *Applied multivariate statistical analysis*. Prentice-Hall, New Jersey, 2007.

[8] Klecka, W. R.: *Discriminant analysis*. Sage Publications, Newbury Park, 1980.

[9] McLachlan, G. J.: *Discriminant analysis and statistical pattern recognition*. Wiley, Hoboken (New Jersey), 2004.

[10] Scott, D. W.: *Multivariate density estimation: Theory, practice, and visualization*. Wiley, New York, 1992.

[11] Silverman, B. W.: *Density estimation for statistics and data analysis*. Chapman & Hall, London, 1986.

[12] Simonoff, J. S.: *Smoothing methods in statistics*. Springer, New York, 1996

[13] Vopatová, K.: *Bandwidth selection for kernel density estimates*, PhD thesis (in Czech). Masaryk University, Brno, 2011.

[14] Wand, M. P., Jones, M. C.: *Kernel smoothing*. Chapman & Hall, London, 1995.

Analytical microeconomics as a search tool for the stock optimization

Simona Hašková[1], Robert Zeman[2]

Abstract. In the paper we look into the inventory stock optimization problem within which we focus on conditions of uncertainty. We compare two approaches: 1. we concentrate on the exact approach of a mathematician, who derives his conclusions from a minimum of factual findings and who mostly leans on the formal side of his model; 2. we deal with the same thing applying the "case-based reasoning" approach, which is based on the detailed knowledge and which is understandable to a manager. This comparison enables us to reveal a number of interesting relations and differences. The cooperation of the two approaches enables a deeper insight into the principle of the optimization algorithm. This algorithm is then applied to the practical case study solved under conditions of uncertainty.

Keywords: microeconomic analyses, theoretical approach to optimization, "case-based reasoning" approach to optimization, uncertainty conditions.

JEL Classification: C51, D87

AMS Classification: 91B32, 91 B70, 91B82

1 Introduction

Basically, there are two approaches to the rational solution of specific given tasks: one of them (the exact approach) proceeds "from above", the second "from bellow". The procedure "from above" is based on the model for the solution of a general problem, which is then modified to "suit" the particular task (it is adapted to it) by appropriate choice of free parameters. If we, for instance, should determine the unknown length x of a tread of the rectangular shape of the width c obliquely laid in a square room of the side a ($a > c$) so that each of its corners touch the different walls of the room (see [8]) we can proceed from the pair of formulas $x^4 + (-b^2 - a^2 - 2c^2)x^2 + 4abcx - c^2(a^2 + b^2) + c^4 = 0$, $c < \min\{a, b\}$, which is the model for the solution to the same task for the rectangular room with sides a, b (free parameters). Substituting $a = b$ we get a special model in the form $x^4 - 2(a^2 + c^2)x^2 + 4a^2cx - 2a^2c^2 + c^4 = 0$, $c < a$ valid for a square room, from which we can calculate the length of x from the entered values a and c by means of one of the conventional iterative methods. As we can see, a special model is no simpler than the more general model, neither is its solution any easier.

The procedure „from bellow" (called the "case-based reasoning" approach), by contrast, builds on knowledge of the specifics of a particular task that is adequate to the selected perspective of the problem. Only those of them are chosen that help to find the solution and contribute significantly to its effectiveness [1]. Unlike the approach, in which it is often necessary to ransom a higher level of generality by the "deletion" (or negligence) of the specifics of the particular tasks, the second approach often enables us to find a more acceptable or efficient solution. Regarding the task of the length x of a tread it is possible, for example, to take into account the specificity of a square room (the centre theme of the tread has to lie on the diagonal of the room) and to formulate a much simpler model of the form $x = 2^{1/2} \cdot a - c$ for a square, which in the case of a rectangle does not apply. This model cannot be formally deduced from its complex version, even though (as we can see after direct substitution of the last to the previous condition) both models correspond to each other.

The literature on inventory theory, which is developed by theorists (mostly mathematicians), prefer the approach "from above". It is based on the general optimization principles within which the abstract framework is formulated (see e.g. [4] in the case of inventory optimization problem and e.g. [21], [22] in the case of other optimization problems). This is then fulfilled by the mathematical description of ideas about the possible functioning of a warehouse.

Another difficulty lies in the fact that a manager is not usually trained in mathematics or logic, so the application of abstract and substantively empty laws of logic and probability (the principles of "rational reasoning") at a general level is often not "user friendly" for him (the error may not be in the person, but in the way in which the case is presented - see [6]). Moreover, as a number of cognitive psychologists argue (see e.g. [7], [13] and

[1] Institute of Technology and Business, Department of Economics and Management, Okružní 517/10, 370 01 České Budějovice, CZ, e-mail: haskovas@post.cz

[2] Institute of Technology and Business, Department of Economics and Management, Okružní 517/10, 370 01 České Budějovice, CZ, e-mail: zeman@mail.vstecb.cz

[20]), the manager is in this respect usually restrained by the „limited rationality“. It is much easier for him to rely on knowledge of the principles of analytical microeconomics, build on the specifics of a particular problem and the ability to perform elementary managerial calculations. This in the case of the “case-based reasoning” approach (the essence of which is explained in [8]) in combination with a certain amount of proficiency in the “art of solving problems” is often sufficient as shown in practical case studies [15], [16] , [17] and [18].

This paper is a continuation of [9], which mainly deals with the derivation of the inventory optimization model. Here we base on the therein derived relations when solving the specific inventory problem the solution of which is constructed step by step within the case-study. This we consider to be the main contribution both to the theory and practice. We demonstrate an application of the “case-based reasoning” approach to resolve the frequently occurring inventory task, which might be difficult to solve for the “usual” manager within the “from above” procedure.

2 The situational study as a database for inventory case study

The data of the inventory task are drawn from the empirical situational study provided by an electronic retailer who sells, among other things, a certain brand of washing machines. Based on them the following probability distributions of the daily demand values for washing machines (*JP*) and of the supplier delivery times (*DD*) (i.e. the time from order submission to the washing machines delivery to the store) shown in Tab. 1 and Tab. 2 were obtained from the analytical evidence of financial accounting over several years (daily sales plus records of customer dissatisfaction due to the depletion of stocks, orders and delivery notes as also demonstrated in [2]):

daily demand *JP* (piece)	**probability *JP***
0	0.11
1	0.30
2	0.27
3	0.18
4	0.09
5	0.04
6	0.01

Table 1 The probability distribution of the daily demand (*JP*)

delivery time *DD* (in days)	**probability *DD***
1	0.3
2	0.5
3	0.2

Table 2 The probability distribution of delivery times (*DD*)

Furthermore, the situational study states that the washing machines are bought from a supplier at the wholesale price of 10000 CZK and sells them for a retail price of 12000 CZK per unit. The failure to sell one machine caused by the loss of a customer means the loss of 2000 CZK for the retailer (the lost unit margin). Depending on the level of stock the store orders and imports to its storehouse 50 washing machines almost every month (i.e. on average 12 times per year). The short-term loan interest rate of this store is 10 % p.a.; by the storage of one machine the store loses 1000 CZK per year on interest. All costs associated with the operation of warehouse management (including warehouse staff wages) are fixed and do not depend on the number of stored washing machines.

From it we assign:

- X_{opt} = 50 to the volume of one additional supply of washing machines,
- H = 1000 CZK to the variable part of the annual costs associated with the storage of a washing machine,
- Z = 2000 CZK to the loss due to customer dissatisfaction,
- $E[D]$ = 600 pieces to the amount of the expected annual orders of washing machines to the store.

The task is: To determine the optimal reorder level (i.e. stock level at which it is necessary to place an order for an additional supply (HD_{opt})), provided that the quantity demanded each day is independent and the store intends to maximize profit.

3 The methodology: The exact approach versus the "case-based reasoning" approach to the solution

Within the exact approach (the procedure "from above") the data stated above is to be substituted into the appropriate model together with the data of both tables characterizing the corresponding random variables (daily demanded quantity (*JP*) and delivery time of an additional supply (*DD*)). But in which one and how?

In technical publications aimed at university-educated managers (see e.g. [3], [10], [11], [14] or [19]) we can find dozens of models differentiating between in the single-product, multi-product, deterministic, stochastic, static or dynamic ones, discrete or continuous, with a stationary or non-stationary demand, etc. In the category of discrete dynamic stochastic models (in terms of inventory management by so-called Q-systems), to which our case belongs, the manager encounters a general formulation of aggregate demand (not in the economic sense, but in terms of a gradually accumulated demanded quantity) described by a stochastic process defined on the considered period of warehouse functioning. The manager is then forced to face the problems associated with it [14]: he/she has to verify whether the considered decomposition of aggregate demand on a one-day demand satisfies the conditions of the central limit theorem; he/she has to consider in what way to properly take into account the random delivery time when the models consider it as a constant or dependent on the ordering cycles; eventually he/she is forced to approximate the available empirical distribution of *JP* and *DD* in models by the required theoretical distributions ($N(\mu,\sigma^2)$, algebraic polynomials, etc.).

In the face of these problems the manager often recalls what is, in practice, confirmed by a number of studies and what is in [12] summarized in the explicit conclusions: "An algorithm that you have written on a piece of paper is often as good as the statistical formula with multiple regression." or "A complex statistical algorithm does not add any value to the quality of a prediction (or only a minimal one).", etc. This motivates the manager to take into account only significant specifics of the task, to analyze the problem by means of the tools that he/she understands and to utilize the procedures that he/she can justify. From this idea it is only a short step to the "case-based reasoning" approach.

The application of the "case-based reasoning" approach on the problem of inventory optimization is elaborated in detail in [9] where we focused on the construction of the derivative procedure leading to optimization algorithm in a discrete and continuous case.

From it, for the discrete case it turned out that to answer to the problem one of the basic questions "when to reorder?" (which we ask in the case study demonstrated in part 4 and which explicitly requires to find out optimal reorder level) it is sufficient to know the readily ascertainable value of the parameter $\Phi = H / (Z \cdot n)$ where $n = E[D] / X_{opt}$ is the annual number of additional supplies and X_{opt} is the optimal size of an additional supply. The key idea of it lies in proceeding from the maximal value of the demand during the delivery time ($PBDD_{max}$), which corresponds to the maximal reorder level (HD_{max}), backwards. Within this process two kinds of marginal values are compared: the marginal saving of the annual storage costs associated with safety stock costs ($H \cdot PZ - H \cdot (PZ - 1) = H$, where *PZ* stands for the variable safety stock level) and marginal expected annual costs associated with the shortage of stock ($\Delta E[ZNZ(HD)] = E[ZNZ(HD - 1)] - E[ZNZ(HD)]$, where *ZNZ* is the annual loss resulting from the lack of stock). The optimization algorithm determines the HD_{opt} as that one that accommodates the condition $\Delta E[ZNZ(HD)] \geq H > \Delta E[ZNZ(HD + 1)]$. This procedure is captured in Fig. 1.

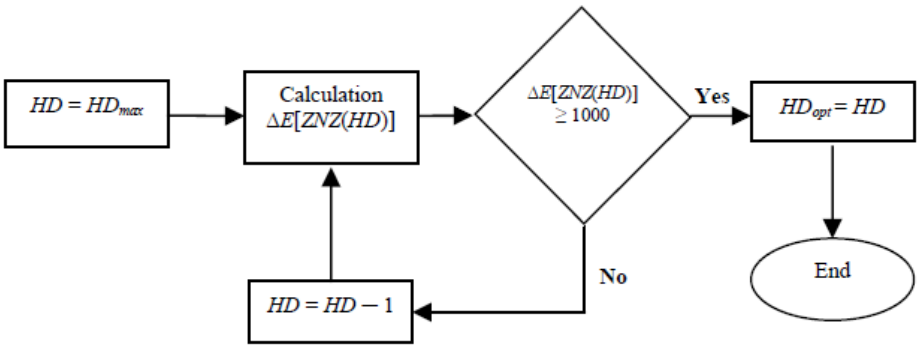


Figure 1 Flow-chart of optimization algorithm with H = 1000 CZK

The quantification of $\Delta E[ZNZ(HD)]$ for different values of *HD* leans on the probability distribution of the demand during the delivery time (*PBDD*). It can be derived from the known probability distributions of random variables *JP* and *DD*. The distribution *PBDD* is in [9] defined as follows:

Given a set $\Omega = \{0, 1, \dots, pbdd_{max}\}$ of integer values of the variable *PBDD* and the projection $p: \Omega \rightarrow \langle 0, 1\rangle$ such that for the sum of $p(i)$ over all $i \in \Omega$ applies $\Sigma p(i) = 1$. The projection p defining the set of pairs $\{(i, p(i)): i \in \Omega\}$ is called probability distribution of a discrete random variable *PBDD* where i is the current value of *pbdd* and $p(i)$ its probability occurrence.

The searched HD_{opt} (optimal reorder level) is then the $HD \in \Omega$, which fulfills the optimization condition derived in [9]:

$$p(HD) + \sum p(i) \geq \Phi > \sum p(i) \tag{1}$$

where $\Phi = H \cdot X_{opt} / (Z \cdot E[D]) = H / (Z \cdot n)$ is a constant; in both summations it is summed over all $i > HD$.

This "from bellow" procedure respects all the significant characters of the task and builds on them. In the next part it is applied to the solution of the inventory problem the input data of which are summarized in part 2.

4 The case study - solution by the "case-based reasoning" approach

The answer to the question "When to reorder an additional supply?" posed in part 2 is to be found by following procedure within which the efficient way of calculation the probability distribution *PBDD* is introduced.

From the given data we calculate $\mathbf{\Phi} = 1000 / (2000 \cdot 12) = 1 / 24 =$ **0.0417**. The field $\Omega = \{0, 1, \dots, 18\}$ of variable values *PBDD* is the disjoint union of three disjunctive components: $\Omega_1 = \{0, 1, \dots, 6\}$, $\Omega_2 = \{7, 8, \dots, 12\}$ and $\Omega_3 = \{13, \dots, 18\}$; $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$.

The value of $pbdd = i \in \Omega_3$ is reached by the simultaneous occurrence of two conditioned phenomena: $dd = 3$ ($p(dd = 3) = 0.2$) and $i = jp_1 + jp_2 + jp_3$, where jp_k is the demanded quantity in the *k-th* day of the delivery time, $k = 1, 2, 3$. Let $\omega_3(i) \subset \Omega_1^3$ be the set of all triples (jp_1,jp_2,jp_3), satisfying the condition $jp_1 + jp_2 + jp_3 = i$. Given the assumption of the mutual independence of the elements of the trio we can write the probability of its occurrence conditioned by $dd = 3$ as $p_3((jp_1,jp_2,jp_3)) = p(jp_1) \cdot p(jp_2) \cdot p(jp_3)$. Various triads are incompatible phenomena, and therefore their probabilities of occurrence can be added. We denote as $p_3(i)$ the value of the expression $\Sigma p_3((jp_1,jp_2,jp_3))$, in which is summed over all $(jp_1,jp_2,jp_3) \in \omega_3(i)$. Then $p(pbdd = i) = p(dd = 3) \cdot p_3(i) = 0.2 \cdot p_3(i)$.

For instance, the calculation of $p(pbdd = 13)$ can proceed as follows: The condition $jp_1 + jp_2 + jp_3 = 13$ is met by each trio consisting of the numbers: 2,5,6; 3,4,6; 1,6,6; 3,5,5; 4,4,5.
$p_3((2,5,6)) = 0.27 \cdot 0.04 \cdot 0.01 = 0.000108$
$p_3((3,4,6)) = 0.18 \cdot 0.09 \cdot 0.01 = 0.000162$
$p_3((1,6,6)) = 0.30 \cdot 0.01 \cdot 0.01 = 0.00003$
$p_3((3,5,5)) = 0.18 \cdot 0.04 \cdot 0.04 = 0.000288$
$p_3((4,4,5)) = 0.09 \cdot 0.09 \cdot 0.04 = 0.000324$

Each of the first two triads has six equally probable permutations. The remaining trios have three equally probable permutations. Therefore $p_3(13) = 6 \cdot 0.000108 + 6 \cdot 0.000162 + 3 \cdot 0.00003 + 3 \cdot 0.000288 + 3 \cdot 0.000324 = 0.003546$ and therefore $\mathbf{p(pbdd = 13)} = 0.2 \cdot 0.003546 =$ **0.0007092**.

From the course of the probability distribution of variable values *JP* in Tab. 1 it is obvious that $p(pbdd = i)$ with the increase of $i \in \Omega_3$ decreases. Therefore, the left part $p(HD) + \Sigma p(i) \geq \Phi$ of the condition (1) in the set Ω_3 cannot be fulfilled.

HD_{opt} is thus to be found in the remainder of $\Omega - \Omega_3 = \Omega_1 \cup \Omega_2$ of the set Ω. We start with the set Ω_2 and, as in the case of the set Ω_3, we try to clarify the position of HD_{opt} based on the knowledge of $p(pbdd = 7)$, i. e. from the knowledge of the probability of the occurrence of its smallest element.

The value of $i \in \Omega_2$ occurs when $dd = 3$ and $i = jp_1 + jp_2 + jp_3$ or when $dd = 2$ a $i = jp_1 + jp_2$. Both options are mutually exclusive, so it applies that $p(pbdd = i) = p(dd = 3) \cdot p_3(i) + p(dd = 2) \cdot p_2(i) = 0.2 \cdot p_3(i) + 0.5 \cdot p_2(i)$, where $p_3(i)$ is already known and $p_2(i) = \Sigma p_2((jp_1,jp_2))$, summing over all $(jp_1,jp_2) \in \omega_2(i) \subset \Omega_1^2$ satisfying the condition $jp_1 + jp_2 = i$.

The condition $jp_1 + jp_2 + jp_3 = 7$ meets each trio consisting of the numbers: 0,1,6; 0,2,5; 0,3,4; 1,2,4; 1,1,5; 2,2,3; 1,3,3. Therefore, $p_3(7) = 6 \cdot (0.00033 + 0.001188 + 0.001782 + 0.00729) + 3 \cdot (0.0036 + 0.013122 +$

0.00972) = 0.142866. The condition $jp_1 + jp_2 = 7$ meets each pair consisting of numbers: 1,6; 2,5; 3,4. Therefore, $p_2(7) = 2 \cdot (0.003 + 0.0108 + 0.0162) = 0.06$. Thus $\boldsymbol{p(pbdd = 7)} = 0.2 \cdot 0.142866 + 0.5 \cdot 0.06 = \mathbf{0.0585732}$.

Given that $p(pbdd = 7) > \Phi$, the right part $\Phi > \Sigma p(i)$ of the condition (1) cannot be fulfilled in the set Ω_1. Thus $HD_{opt} \in \Omega_2$, and therefore we do not need the values of $p(pbdd = i)$ for $i \in \Omega_1$ for the solution to the tasks. If we still find all elements of the distribution $\{(i, p(i)): i \in \Omega\}$ of variable *PBDD* in the manner outlined, we get the following Tab. 3:

pbdd	0	1	2	3	4	5	6	7	8	9	10
p(pbdd) %	4	12	16	17	15	12	11	6	3	2	1

Table 3 The fragment of the probabilistic distribution of the variable *PBDD*

In Tab. 3, the probabilities are expressed in percentage rates and then rounded to the nearest integer values. Values of $p(pbdd)$ % for $pbdd > 10$ are negligible (rounded to 0), which is the reason why this part of the table is dropped. The sum of the bottom row gives 99 %, which means that inaccuracies in the table are not significant. If we express the parameter Φ as a percentage, we get Φ % = 4.17. From the condition (1) it follows that $\boldsymbol{HD_{opt} = 8}$.

For $E[PBDD]$ it applies approximately:

$\boldsymbol{E[PBDD]} = 0.12 + 2 \cdot 0.16 + 3 \cdot 0.17 + 4 \cdot 0.15 + 5 \cdot 0.12 + 6 \cdot 0.11 + 7 \cdot 0.06 + 8 \cdot 0.03 + 9 \cdot 0.02 + 10 \cdot 0.01 = \mathbf{3.75}$, from which we get the safety stock level $\boldsymbol{PZ} = HD_{opt} - E[PBDD] = \mathbf{4.25}$.

5 Conclusion

The exact approach of mathematicians dealing with the field of inventory theory produces more sophisticated and complex models, which often extend beyond the constraints of the limited rationality of managers. An example of this might be found in [4], in which a deterministic single-product inventory model with two separate warehouses and variable demand is presented, and it is formulated as a nonlinear programming solution by genetic algorithm. However, as we know "expensive" does not always mean "better" and "complex" does not always mean "correct" (see e.g. [5]). The information in [4] that "all the simulations have been conducted on a PC with Intel Core-2 duo processor, 1 GB RAM and 2.0 GHz speed" does not increase the trust of a manager in a mechanism which he does not understand.

Therefore, it is important to highlight the existence of other approaches to be used by management, particularly those relying on knowledge that the manager is familiar with and is able to justify within the solution. This shall include: knowledge of the specifics of the problem that the manager solves (experience), basic knowledge of microeconomic analysis including erudition in the field of managerial calculations related to it, knowledge of probability theory and statistics, and last but not least, common sense and the ability to utilize it properly. All the above mentioned is the basis of the "case-based reasoning" approach. This approach has been used to resolve the inventory optimization problem shown in a case study. As follows from the solution procedure, the correct result of the task can be estimated from the known value of the parameter Φ and from the several values of distribution *PBDD*, while the requirements for the accuracy of these values are not great; they are fulfilled.

References

[1] Ackoff, R.L., Addison, H.J., Carey, A.: *Systems thinking for curious managers: With 40 new management F-laws*. HJ Addison, A. Carey, 2010.

[2] Akcay, A., Biller, B., Tayur S.: Improved inventory targets in the presence of limited historical data. In: *Manufacturing & Service Operations Management*, Vol.13, (2011), 297-309.

[3] Ang, M., Song, J.S., Wang, M., Zhang, H.: On properties of discrete (*r*, *q*) and (*s*, *T*) inventory systems. In: *European Journal of Operational Research*, Vol. 229, Iss. 1, (2013), 95–105.

[4] Bhunia, A. K., Pal, P., Chattopadhyay, S., and Medya, B. K.: An inventory model of two-warehouse system with variable demand dependent on instantaneous displayed stock and marketing decisions via hybrid RC-GA. In *International Journal of Industrial Engineering Computations*, (2011), 351–368.

[5] Bono, E.: *Simplicity*. Penguin Books Ltd, 1990.

[6] Gigerenzer, G.: Ecological intelligence: An adaptation for frequencies. In: *The Evolution of Mind (D. D. Cummins & C. Allen)*, (1998), 9-29.

[7] Gigerenzer, G., and Gaissmaier, W.: Heuristic Decision Making. In: *Annual Review of Psychology*, Vol. 62, (2011), 451-482.

[8] Hašková, S. and Kolář, P.: The Limits of Relativism in Managerial Thinking. In: *Proceedings of 12th European Conference on Research Methodology for Business and Management Studies* (Ramos, I., Mesquita, A.). Univ Minho, Guimaraes, Portugal, 2013, 153-161.
[9] Hašková, S., Kolář, P.: Microeconomic view on the inventory optimization problem. In: *Proceedings of the 31st International Conference on Mathematical Methods in Economics „MME 2013"* (Hana Vojáčková). College of Polytechnics Jihlava, 2013, 232-237.
[10] Jiroušek, R.: Brief Introduction to Probabilistic Compositional Models. In: *Uncertainty Analysis in Econometrics with Applications, Advances in Intelligent Systems and Computing*, Vol. 200, (2013), 49-60.
[11] Jiroušek, R., Shenoy, P.P.: Compositional models in valuation-based systems. In: *International Journal of Approximate Reasoning*, Vol. 55, Iss. 1, Part: 3, (2014), 277-293.
[12] Kahneman, D.: A perspective on judgment and choice: Mapping bounded rationality. In: *The American Economic Review*, Vol. 58(9), (2003), 697-720.
[13] Kahneman, D., and Tversky, A.: On the psychology of prediction. In: *Psychological Review*, (1973), 237-251.
[14] Lukáš, L.: *Pravděpodobnostní modely v managementu*. 2. Díl, Academica, 2012.
[15] Maroušek, J., Hašková, S., Zeman, R. & Vaníčková, R.: Managerial Preferences in Relation to Financial Indicators Regarding the Mitigation of Global Change. In: *Science and Engineering Ethics*, (2014), 1-5.
[16] Maroušek, J.: Study on Agriculture Decision-Makers Behavior on Sustainable Energy Utilization. In: *Journal of Agricultural & Environmental Ethics*, Vol. 26, Iss. 3, (2013), 679-689.
[17] Maroušek, J.: Economically oriented process optimization in waste management. *In Environmental Science and Pollution Research*, (2014), 1-3.
[18] Maroušek, J., Hašková, S., Zeman, R., Váchal , J., Vaníčková, R.: Assessing the implications of EU subsidypolicy on renewable energy in Czech Republic. In: *Clean Technologies and Environmental Policy*, (2014), 1-8.
[19] Ross, S.M.: *Applied probability models with optimization applications*. Holden-Day, Inc., San Francisco, 1970.
[20] Simon, H. A.: *Models of Bounded Rationality: Empirically Grounded Economic Reason*. Massachusetts Institute of Technology, 1982.
[21] Volf, P.: On problem of optimization under incomplete information. In: *Proceedings of 30th International Conference on Mathematical Methods in Economics* (Ramik, J., Stavarek, D.). Silesian Univ Opava, Sch Business Adm Karvina, Czech Soc Operat Res; Czech Econometr Soc., Karvina, 2012, 968-973.
[22] Volf, P.: A stochastic model of progression of athletic records. In: *IMA Journal of Managerment Mathemathic*, Vol. 22, Iss. 2, (2011), 157-169.

Title of the contribution Gender Pay Gap in Different Sectors of Czech Economy

Veronika Hedija[1]

Abstract. The study is devoted the wage differences between men and women in the Czech Republic. It aims to determine whether and to what extent the unexplained gender wage gap varies in the different sectors of the Czech economy using classification NACE Rev. 2 and if so, identify the possibly causes of these differences.

We use EU-SILC data for 2010. Firstly, we estimate average treatment effect on the treated for individual sectors of Czech economy which enable identify the unexplained part of gender pay gap. To identify the possibly causes of observed variability in ATT, we regress ATT (calculated for every category according to classification NACE Rev.2) by ownership (public or private sector), proportion of women in management, proportion of women in sector and proportion of small firms.

We conclude that the unexplained gender pay gap vary in individual sectors of Czech economy. The estimated ATT ranges from -0.004 (NACE L-N) to -0.352 (NACE G). The most important factor explaining these differences is ownership as expected. The significant factor is the proportion of women in management too. On the other hand, the role of proportion of women in sector and dominated firm size is negligible.

Keywords: average treatment effect on the treated, gender pay gap, wage differences, labor market.

JEL Classification: J16, J24, J71
AMS Classification: 91G70

1 Introduction

The fact that women earn on average less than men is well known and accepted in economic literature. Many studies are devoted the issue of wage differences between men and women and identification of their causes. To identify the causes of gender wage differences, these studies mainly use Oaxaca-Blinder decomposition or its modification which enable to separate the part of gender pay gap which could be explained by differences in known observed personal and firm characteristics of men and women and the part which could not be explained by these and which is often called as "remuneration effect" or "effect of discrimination". The results of these studies depend on used dataset, number of explanatory variables and applied method of decomposition (for more detail Beblo et al. [1]). However, all of these conclude that part of gender wage differences have stayed unexplained.

The raw gender pay gaps and their unexplained part vary by country. Christofides, Polycarpou and Vrachimis [5] use EU-SILC data and estimate the unexplained part of gender pay gap for 26 European countries including the Czech Republic using Oaxaca-Ransom decomposition. The raw gender pay gap amounted 0.27 in the Czech Republic and it was among the highest in the Europe. The different characteristics of men and women enabled to explain approximately one quarter of raw gender pay gap, the rest remained unexplained. Firm characteristics are important factor explaining the existing gender wage difference in the Czech Republic, personal characteristics play only minor role (Jurajda [6], Mysíková [7], Hedija [2]). Jurajda [6] analysed the gender pay gap in Czech and Slovak private and public sector. He concluded that the unexplained gender pay gaps differed dramatically between both sectors. In the private sector it represented more than 60 percent of the raw gender pay gap in comparison to the public sector, where it amounted approximately 40 percent of the gender pay gap. These differences may be explained by different wage setting mechanism in private and public sector and easily implementation of anti-discrimination policies in public sector.

According to the Eurostat data, the raw gender pay gap varies not only in private and public sector but also in the individual sectors of Czech economy. The part of these differences could be probably explained by different average firm characteristics of men and women in individual sectors. The questions are: what part of gender

[1] Vysoká škola polytechnická Jihlava, katedra ekonomických studií, Tolstého 16, 586 01 Jihlava, veronika.hedija@vspj.cz

wage differences remains unexplained in individual sector, how large are the differences between sectors and what is the cause of existing disparity, if any. According to my knowledge the study dealing with this issue with the application of the Czech Republic does not exist.

The aim of this study is to determine whether and to what extent the unexplained gender wage gap varies in the different sectors of the Czech economy using classification NACE Rev. 2 and if so, identify the possibly causes of these differences.

The first section of the paper describes used methods and data set. To estimate the unexplained part of the gender pay gap we calculate the average treatment effect on the treated (ATT) for individual sectors of Czech economy. We identify the significant variability in unexplained gender pay gap calculated for individual sectors. We attempt to explain this variability using linear regression model with ATT as dependent variable. The final section of the paper summarizes the obtained results.

2 Data

We use data from the European Union Statistics on Income and Living Conditions (EU-SILC). The EU-SILC covers multidimensional micro-data on income, poverty, social exclusion and living conditions.[2] We use cross-sectional data for 2010 which comes from EU-SILC 2011 and covers data of 30 European countries. We selected data for the Czech Republic.

	men	**wo men**	**% of wo men**		**men**	**wo men**	**% of wo men**
Age	42.1	42.9	-	Sector (NACE Rev. 2)			
Education (ISCED-97)				Agriculture, forestry …(A)	0.06	0.02	21.3
lower secondary	0.04	0.06	54.9	Mining, manufacturing…(B-E)	0.42	0.24	31.7
upper secondary	0.79	0.75	43.5	Construction (F)	0.11	0.02	12.9
post-secondary non tertiary	0.01	0.02	61.9	Wholesale, retail trade… (G)	0.08	0.14	58.7
tertiary education	0.17	0.17	44.8	Transportation and storage (H)	0.10	0.05	28.9
Sickness	0.08	0.13	56.9	Accommodation and food …(I)	0.02	0.05	67.0
Partnership (partner)	0.70	0.69	44.4	Information and communication (J)	0.03	0.01	21.3
Children (dependant)	0.67	0.65	44.0	Financial and insurance… (K)	0.02	0.04	61.9
Occupation (ISCO-88)				Real estate, professional, scientific activities …(L-N)	0.05	0.05	44.8
Legislators, senior officers and managers	0.06	0.04	35.1	Public administration and defence …(O)	0.07	0.09	51.0
Professionals	0.09	0.11	49.8	Education (P)	0.03	0.12	76.4
Tech. and associate professionals	0.19	0.30	56.2	Human health, social activities (Q)	0.02	0.13	84.1
Clerks	0.04	0.17	77.5	Art, entertainment …(R-U)	0.06	0.03	28.9
Service and sales workers	0.07	0.16	65.0	Firm size (>10 employees)	0.84	0.78	43.0
Skilled agricultural, fishery workers	0.01	0.01	44.8	Contract (temporary)	0.07	0.08	48.1
Craft and related trades workers	0.28	0.06	14.8	Managerial position	0.23	0.16	36.1
Plant and machine operators …	0.20	0.08	24.5	Work experience	22.0	21.0	-
Elementary occupations	0.04	0.07	58.7	Hourly gross wage	1.72	1.48	-
Armed forces	0.01	0.00	0.0	N	3494	2834	39.7

Table 1 Average of characteristics of men and women and proportion of women

EU-SILC data do not contain information on hourly wages, it was necessary to narrowed the sample to be able calculate these using available data. We narrowed the reference population sample to persons who were employees in the reference period, worked all twelve months in full-time job, had no other jobs and earned an income. We excluded the self-employed because the wage differences among self-employed men and women say nothing of potential wage discrimination against women. Their earnings are closely linked to the realized or declared profit and are not dependent on the decision of another person. We use data on selected personal and firm characteristics of employee: education level (highest attained education level according to ISCED-97),

[2] It provides two types of data: cross-sectional and longitudinal. The former includes data to the given time or a certain time period, the second contains individual-level changes over time. The reference populations in EU-SILC are all private households and their current members residing in the territory of the countries at the time of data collection. The personal data comes from household members aged sixteen and plus.

sickness (temporary inability to work due to sickness in income reference year), partnership (having partner in common household), dependent children (having at least one economic inactive child under 24 years), occupation (according to the classification ISCO-88), sector (economic activity using classification NACE Rev.2), firm size (less than 11 employees and 11 and more employees), contract (having work contract of limited duration), managerial position (having formal responsibility for supervising a group of other employees), work experience (number of year spent in paid work) and hourly gross wage. Hourly gross wage is calculated as employee cash and non-cash incomes per year divided by the number of hours usually work per year (including overtime).

Brief overview of the data set is presented in Table 1. It shows average of observed characteristics of men and women in the sample and proportion of women in individual categories.

3 Methods

To identify the differences in unexplained part of gender pay gap in individual sectors and find out the potential causes of its variability, if any, we firstly estimate the unexplained part of gender pay gap in individual sectors of Czech economy and then use regression model with unexplained part of gender pay gap as dependent variable.

To establish the unexplained part of the gender pay gap we estimate the average treatment effect on the treated which reflects the part of raw gender pay gap which could not be explained by differences in known observed characteristics of men and women in the sample and which could be the result of wage discrimination against women. We counted the ATT for individual sectors of Czech economy according to NACE Rev. 2. Average treatment effect on the treated is average benefit resulting from being treated. In our case ATT is the mean effect for women in the form of lower wage resulting from being women. We use this form for the average treatment effect on the treated calculation

$$ATT = E\left(y_i(1) - y_i(0) \middle| T_i = 1\right). \tag{1}$$

Where T is the binary treatment indicator, T = 1 denotes treatment and T = 0 otherwise, y(1) is the potential outcome with treatment and y(0) is the potential outcome without treatment. In our case, be treated means be the women. We can rewrite the ATT as

$$ATT = E\left(y_i(1) \middle| T_i = 1\right) - E\left(y_i(0) \middle| T_i = 1\right). \tag{2}$$

Where ATT represents the gender pay gap which cannot be explained by the different characteristics of men and women. The term $E(y_i(1)|T_i=1)$ is the sample average of logarithm of gross wage of women and the term $E(y_i(0)|T_i=1)$ is the sample average of logarithm of gross wage of women, if they were men. From our sample, we know the first term on the right-side of the equation 2, the sample average of logarithm of hourly gross wage of women. The second term, the average of logarithm of women hourly gross wage if they were men, we must estimate someway. There are more ways to estimate this. For more details see Wooldridge [10] and Ho et al. [4]. We estimate this using the regression model.

First, we estimate the coefficients of the wage function of men

$$\left(y_i \middle| T_i = 0\right) = \beta_0 . X_i + u_i . \tag{3}$$

Where y_i is the logarithm of the male gross hourly wage, β_0 is the vector of the coefficients of the wage function, X is the vector of the chosen observed characteristics of men and u is a disturbance term. As explanatory variables we use education level, work experience, squared work experience, sickness, partnership, dependent children, occupation, firm size, contract and managerial position.

Then we use the estimated coefficients of the male wage function to compute the average of logarithm of gross hourly wage of women, if they were men.

$$E\left(y_i(0) \middle| T_i = 1\right) = E(\beta_0 . X_i). \tag{4}$$

Where $E(\beta_0.X_i)$ is the mean of the predicted wages (logarithm of gross hourly wage) of every woman in the sample. Finally, we estimate the average treatment effect on the treated as the difference between the average of logarithm of gross hourly wage of women and the average of predicted values of wages computed from male wage function.

$$ATT = E\left(y_i(1) \middle| T_i = 1\right) - E\left(\beta_0 . X_i \middle| T_i = 1\right). \tag{5}$$

Then we construct the linear regression model where we use ATT calculated for individual sectors using equation 5 as dependent variable.

$$ATT = \alpha + \beta_1.public_i + \beta_2.women_i + \beta_3.femalemanagers_i + \beta_4.smallfirms_i + u_i \quad (6)$$

Where i denotes the sector, $public_i$ is the dummy for public sector, $women_i$ is share of women in sector i, $femalemanagers_i$ denotes share of female managers in sector i and $smallfirms_i$ is the proportion of small firm in sector i.

We use dummy for public sector, proportion of female managers, proportion of women in sector and proportion of small firms as explanatory variables. Here is a description of the variables and arguments for their inclusion. Jurajda [6] states that unexplained gender pay gap is lower in public sector as compared with private sector. The reason may be different wage setting mechanism in both sector and easily implementation of anti-discrimination policy in public sector which is regulated. We use dummy variable for public sector, this denotes the sectors with dominant share of public sector (NACE O, P and Q). Hedija [3] confirms that the presence of female managers lead to a decrease in the gender pay gap. These conclusions are in accordance with social identity theory. This theory states that individuals tend to favour members of their own group from other group members (Tajfel [9]; Tajfel and Turner [10]). Hence, women in managerial positions, that can affect the wage of their subordinates, are likely to evaluate female employees better than male employees. We use the proportion of female managers as other control variable. It is calculated as the share of female managers on all managers using classification of occupation ISCO-88, where managerial employees are corporate managers (ISCO code 12) and general managers (ISCO code 13). The unexplained gender pay gap may be affected by the proportion of women in the sector. If the proportion of women is low, the women may be different from other women and can have more male characteristics. They could be part of the male teams and may be perceived as men. This can be applied if pay too. To capture the effect of proportion of women on unexplained gender pay gap, we use this as other control variable in model. The proportion of women is calculated as a share of women on all employees in individual sectors. Proportion of small firms is other factor which could possibly determine the amount of gender pay gap. The wage discrimination of women is more difficult in small firm as compared with their large counterparts. The employees of small businesses know each other better and may also disclose the amount of their wages. To maintain a good working environment and good relationship at work, the imposing the wage discrimination against women is more difficult for employers in these firms. The proportion of small firms is calculated as the share of firms with 10 and less employees. To estimate the coefficients of model we use OLS with robust standard errors.

4 Empirical results

We calculated the ATT using equation 5. Firstly we estimate the wage function of men using equation 3. As the explanatory variables we used education, work experience, work experience squared, sickness, partnership, dependent children, occupation, firm size, contract and managerial position. Then we estimated ATT from equation 5 as the difference between average female wage and female wage if they were men. The results are shown in table 2.

	all	**A**	**B-E**	**F**	**G**	**H**	**I**
ATT	-0.195***	-0.232***	-0.250***	-0.041	-0.352***	-0.160***	-0.199***
	(0.013)	(0.058)	(0.017)	(0.094)	(0.038)	(0.053)	(0.059)
N	6328	219	2169	428	657	479	208
Men	3494	152	1475	371	272	332	72
Women	2834	67	694	57	385	147	136
	J	**K**	**L-N**	**O**	**P**	**Q**	**R-U**
ATT	-0.173*	-0.242***	-0.004	-0.148***	-0.104**	-0.143**	-0.098
	(0.096)	(0.059)	(0.089)	(0.035)	(0.042)	(0.072)	(0.085)
N	134	168	313	502	453	447	151
Men	96	57	166	260	101	68	71
Women	38	111	147	242	352	379	80

***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level, robust standard errors in brackets, A - agriculture, forestry and fishing; B-E - mining and quarrying, manufacturing, electricity, gas, steam and air conditioning supply water supply, sewerage, waste management and remediation activities; F - construction; G - wholesale and retail trade, repair of motor vehicles and motorcycles; H - transportation and storage; I - accommodation and food service activities; J - information and communication; K - financial and insurance activities; L-N - real estate activities, professional, scientific and technical activities,

administrative and support service activities; O - public administration and defence, compulsory social security; P - education; Q - human health and social work activities; R-U - arts, entertainment and recreation, other service activities, activities of households as employers, undifferentiated goods- and services-producing activities of households for own use and activities of extraterritorial organisations and bodies.

Table 2 Average treatment effect on the treated

Average treatment effect on the treated which represents the unexplained part of gender pay gap calculated for all economy reaches -0.195. The women earns about 20 percent less than man on average and this gap cannot be explained by differences in known characteristic of men and women. ATT varies in individual sector of Czech economy. It ranges from -0.004 to -0.352. But the ATT calculated for sectors F, L-N and R-U are not statistically significant. Using only statistically significant estimation, the lowest ATT can be identified in the sectors P, Q and O, thus, in sectors with dominant public sector. On other hand, the highest level of potential wage discrimination against women is in sector G (wholesale and retail trade, repair of motor vehicles and motorcycles), where the ATT amounts -0.352.

To explain existing differences in unexplained gender pay gap among individuals sectors of Czech economy, we regress statistically significant ATT. As explanatory variable we use dummy for public sector, proportion of women and proportion of female managers in individual sectors. We use OLS to estimate the coefficient of regression function. The result shows table 3.

	(1)	**(2)**	**(4)**	**(5)**
Constant	-0.230***	-0.188***	-0.194***	-0.194***
	(0.025)	(0.039)	(0.036)	(0.048)
Public sector	0.098***	0.127**	0.120*	0.119**
	(0.028)	(0.053)	(0.053)	(0.033)
Proportion of women	-	-0.001	-0.002	-0.002**
		(0.001)	(0.001)	(0.001)
Proportion of female managers	-	-	0.117**	0.116**
			(0.034)	(0.029)
Proportion of small firms	-	-	-	-0.006
				(0.244)
R^2	0.439	0.505	0.606	0.606
N	10	10	10	10

***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level, robust standard errors in brackets

Table 3 Model of ATT differences by sector: Evidence from OLS

As expected, the most important factor of explaining the differences in ATT between individual sectors of Czech economy is ownership. In public sector the unexplained gender pay gap is approximately 12 percentage points lower as compared with private sector. The causes could be the different wage setting power in private and public sector where public sector faces state regulation in area of earnings, and stricter application of anti-discrimination legislation in public sector. The effect of female managers on unexplained gender pay gap is relatively significant too. The growing in proportion of female managers by 10 percent let to decrease the ATT by 1.17 percentage points. On the other hand the proportion of women and the proportion of small firm (firm with 10 and less employees) in sector play only negligible role. It is not shown that the gender wage differences tended to increase depending on the number of women in the sector or proportion of small firms. Regression coefficients are very low and in the case of proportion of small firms also statistically insignificant.

5 Conclusion

The aim of this study is to determine whether and to what extent the unexplained gender wage gap varies in the different sectors of the Czech economy using classification NACE Rev. 2 and if so, identify the possibly causes of these differences.

We use EU-SILC data for the Czech Republic in year 2010 and estimate the unexplained part of gender pay gap applying average treatment effect on the treated. We know that the unexplained gender pay gaps vary in individual sector of Czech economy (according classification of economic activity NACE Rev. 2). The calculated differences in ATT are considerable. If we consider only statistically significant estimates of ATT, we can conclude that the lower unexplained gender pay gap is identified in sector of education (NACE P). The ATT amounts -0.104 which means that women earn approximately by 10 percent less than men and this wage differences could not be explained by differences in observed characteristics of men and women. On the other

hand the highest ATT is in sector of wholesale and retail trade, repair of motor vehicles and motorcycles (NACE G), where the ATT totals -0.352. Women earn about 35 percent less as compared with men in this sector and this different could be due to wage discrimination against women.

The possibly causes of existing differences in unexplained gender pay gap between individual sectors could be ownership in the sense of domination of private or public sector, proportion of female managers in sector, predominant firm size and proportion of women in sector. We find out that the ownership and proportion of female managers have a significant effect on explaining in differences in unexplained gender pay gap between individual sectors of Czech economy. In sector with dominant public sector, the unexplained gender pay gap is approximately 12 percentage points lower as compared with private sector. Then, the growing in proportion of female managers by 10 percent decreases the ATT by 1.17 percentage points. These findings are consistent with the conclusions of Jurajda [6] and Hedija [3]. Predominant firm size and proportion of women in sector have not significant factor in explaining the variability of ATT between sectors. A weakness of the study is relatively small dataset which is used. This is why this issue and in particular the effect of the proportion of women and company size on unexplained gender pay gap deserves further research.

References

[1] Beblo, M., Beninger, D., Heinze, A. and Laisney, F.: Methodological Issues Related to the Analysis of Gender Gaps in Employment, Earnings and Career Progression. Project carried out for the European Commission Employment and Social Affairs DG, 2003.

[2] Hedija, V.: The Development of Gender Wage Differences in the Czech Republic and Slovakia. In review process in *Communist and Post-Communist Studies.*

[3] Hedija, V.: The Effect of Female Managers on Gender Wage Differences. Accepted in *Prague Economic Papers*.

[4] Ho, D. E., Imai, K., King, G., Stuart, E.A.: Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* **15** (2007), 199-236.

[5] Christofides, L., Polycarpou, A. and Vrachimis, K.: Gender wage gaps, 'sticky floors' and 'glass ceilings' in Europe. *Labour Economics* **21** (2013), 86–102.

[6] Jurajda, S.: Gender Wage Gap and Segregation in Enterprises and the Public Sector in Late Transition Countries. *Journal of Comparative Economics* **31** (2003), 199-222.

[7] Mysíková, M.: Gender Wage Gap in the Czech Republic and Central European Countries. *Prague Economic Papers* **2012/3** (2012), 328-346.

[8] Tajfel, H., Turner, J. C.: An Integrative Theory of Intergroup Conflict. In: *The Social Psychology of Intergroup Relations* (Austin, W. G. & Worchel, S., eds.). Monterey, CA: Brooks-Cole, 1979, 33-47.

[9] Tajfel, H.: Social Psychology of Intergroup Relations. *Annual Review of Psychology* **33** (1982), 1–39.

[10] Wooldridge, J.: *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press, 2002.

Recursive estimators of GARCH models: Selected problems

Radek Hendrych [1]

Abstract. Financial time series modelling is frequently linked to ARCH and GARCH models, which are obviously estimated by the computationally complex (conditional) maximum likelihood estimation procedure. However, in many practical applications, e.g. in the case of high-frequency data, it is necessary to adopt numerically more efficient techniques to calibrate or control such models. The aim of this contribution is to analyse a two stage recursive estimation procedure suggested for the standard GARCH modelling class. Particularly, a Monte Carlo study is performed to examine behaviour of the supposed method. Although the authors of the recurrent technique have presented its adequacy, the results delivered by simulation experiments are not convincing. Such conclusions clearly indicate a need for some revisions; consequently, main ideas of future research are introduced and commented.

Keywords: GARCH, Kalman formulas, recursive estimation, recursive pseudo-linear regression, simulations.

JEL classification: C58.
AMS classification: 62F10.

1 Introduction

The autoregressive conditional heteroskedasticity model [5] and generalized autoregressive conditional heteroskedasticity model [2] (denoted commonly by the acronyms ARCH and GARCH) are nowadays key concepts of econometric and financial modelling. Since their introduction in 1982 and 1986, respectively, they have been investigated from various theoretical and practical perspectives. Particularly, one could surely mention a broad portfolio of successful empirical applications developed mainly in the context of economic and financial time series [10]. Moreover, the simplest GARCH(1,1) model is regarded as a benchmark. However, the parameters of the GARCH models are indeed rarely estimated recursively; the (conditional) maximum likelihood procedure dominates, compare with [6] or [10]. This originates from several serious issues which are naturally connected with recurrent estimation techniques. On the other hand, on-line calibration delivered by applying recursive estimation formulas can be advantageous. For instance, one can mention high-frequency financial time series data, adaptive control processes, fault detections, checking model stability, etc.

The aim of this contribution is to analyse the recursively defined two stage estimation algorithm for the GARCH models considered in [1]. The presented technique is studied by means of a Monte Carlo study in this paper. The results of simulations indicate that the suggested recursive procedure proposed in [1] is disputable. Apparently, this creates a space for further research so that various alternatives are introduced and clarified.

This paper is organized as follows. Section 2 reviews key properties of the GARCH(1,1) model, since it is the most common modelling concept in the supposed framework. Section 3 introduces and comments the recursive estimation scheme proposed by [1]. Section 4 presents a Monte Carlo simulation experiment verifying the recurrent estimation technique. It is shown that the investigated procedure should be revised. Finally, Section 5 concludes and proposes an outline of further research.

[1] Charles University in Prague, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, Radek.Hendrych@mff.cuni.cz

2 GARCH(1,1) model overview

As was mentioned before, the simplest GARCH(1,1) model predicting volatility changes is perceived as the benchmark model by many econometricians and practitioners [6]. Indisputably, it is the most widely applied version in the GARCH class. Accepting this pragmatic argument, the paper is focused only on studying the GARCH(1,1) model. On the other hand, this restraint can be eliminate and the conclusions can be simply generalized. It should be noted that a relatively extensive body of literature concerning the properties of the GARCH models exists, see e.g. [6] or [10] and the references given therein. In order to avoid eventual misunderstandings, the key characteristics of the discussed GARCH(1,1) model are reviewed below.

Definition 1 (Strong GARCH (1,1) process [6])**.** The process $\{y_t\}_{t\in\mathbb{Z}}$ is GARCH(1, 1) if $\mathsf{E}(y_t|\mathcal{F}_{t-1}) = 0$, $\mathsf{var}(y_t|\mathcal{F}_{t-1}) = \sigma_t^2$, $\sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 + \beta_1\sigma_{t-1}^2$, where ω, α_1, and β_1 are parameters, and $\varepsilon_t = y_t/\sigma_t$ are i. i. d. Further, $\mathcal{F}_t$ denotes the smallest σ-algebra with respect to which y_s is measurable for all $s \le t$.

Apparently, the sufficient conditions for $\sigma_t^2 > 0$ (*a.s.*) are $\omega > 0$, $\alpha_1 \ge 0$, and $\beta_1 \ge 0$. If $\beta_1 = 0$, the model is reduced to the ARCH(1) form. Additionally, necessary and sufficient conditions for y_t being (weakly) stationary are $\omega > 0$, $\alpha_1, \beta_1 \ge 0$, and $\alpha_1 + \beta_1 < 1$. The stationary GARCH(1,1) model has the variance $\sigma^2 = \omega/(1 - \alpha_1 - \beta_1)$. For more details and proofs see [6].

Proposition 1 (ARMA representation of the GARCH(1,1) process [6])**.** *Let y_t be the GARCH(1,1) process with $\mathsf{E}(y_t^4) = c < \infty$. Then, y_t^2 can be expressed as an ARMA$(1, 1)$ process with*

$$y_t^2 = \omega + (\alpha_1 + \beta_1)y_{t-1}^2 + v_t - \beta_1 v_{t-1},$$

where $v_t = \sigma_t^2(\varepsilon_t^2 - 1)$ is white noise.

Therefore, under the corresponding assumptions, y_t^2 can be clearly represented as an invertible AR(∞) or a casual MA(∞) linear process. These forms can be also used for estimation or identification purposes. One can also evaluate the k-step ahead prediction of σ_t^2 which is expressed as

$$\hat{\sigma}_{t+k|t}^2 = \frac{\omega}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^k \left(\sigma_t^2 - \frac{\omega}{1 - \alpha_1 - \beta_1}\right). \tag{1}$$

3 Recursive estimation of GARCH parameters

The authors of [1] tried to develop a two stage recursive estimation scheme appropriate for the standard GARCH models following the ideas introduced in [3]. However, they focused mainly on the derivation of the algorithm and not on its numerical evaluation, which can be a crucial objection.

In particular, the whole computational implementation is based on the following representation of the one-step prediction of y_t^2 (where y_t is the GARCH(1,1) process from Definition 1):

$$\hat{y}_{t|t-1}^2 := \mathsf{E}\left(y_t^2|\mathcal{F}_{t-1}\right) = \mathsf{E}\left(\sigma_t^2\varepsilon_t^2|\mathcal{F}_{t-1}\right) = \sigma_t^2. \tag{2}$$

Moreover, y_t^2 can be concisely expressed as

$$y_t^2 = \hat{y}_{t|t-1}^2 + v_t = \sigma_t^2 + v_t, \tag{3}$$

where $v_t = \sigma_t^2(\varepsilon_t^2 - 1)$ is according to Proposition 1 white noise with $\mathsf{E}(v_t) = 0$ and $\mathsf{var}(v_t) = \mathsf{E}(v_t^2)$.

Applying the recursive pseudo-linear regression estimation scheme, see [7] or [9], one can obtain the first stage of the discussed estimation procedure:

$$\begin{aligned}
\tilde{v}_t &= y_t^2 - \tilde{\varphi}_t^\top \tilde{\theta}_{t-1}, \\
\tilde{\theta}_t &= \tilde{\theta}_{t-1} + \frac{\tilde{P}_{t-1}\tilde{\varphi}_t\tilde{v}_t}{1 + \tilde{\varphi}_t^\top \tilde{P}_{t-1}\tilde{\varphi}_t}, \quad \tilde{\theta}_0 = 0, \\
\tilde{P}_t &= \tilde{P}_{t-1} - \frac{\tilde{P}_{t-1}\tilde{\varphi}_t\tilde{\varphi}_t^\top \tilde{P}_{t-1}}{1 + \tilde{\varphi}_t^\top \tilde{P}_{t-1}\tilde{\varphi}_t}, \quad \tilde{P}_0 = \kappa I,
\end{aligned} \tag{4}$$

where $\tilde{\varphi}_t = \left(1, y_{t-1}^2, \tilde{\varphi}_{t-1}^\top \tilde{\theta}_{t-2}\right)^\top$, $\tilde{\theta}_t = (\omega, \alpha_1, \beta_1)^\top$, and κ is a large number, e.g. $\kappa = 10^5$.

Clearly, the recurrent system (4) corresponds to the Kalman filtering formulas associated with the linear discrete-time state space model (compare with [4]):

$$\begin{aligned} \theta_t &= \theta_{t-1}, & (5)\\ y_t^2 &= \varphi_t^\top \theta_{t-1} + v_t, \quad v_t \sim WN\left(0, \mathsf{E}(v_t^2)\right), & (6) \end{aligned}$$

where one puts $\theta_t = (\omega, \alpha_1, \beta_1)^\top$ and $\varphi_t = (1, y_{t-1}^2, \varphi_{t-1}^\top \theta_{t-2})^\top$ similarly as before. The formula (5) describes evolution of unobservable states, i.e. the parameters of the model. The signal equation (6) coincides with the representation given by (3)

The first step obviously does not explicitly take into account the heteroskedasticity of the innovation term v_t. Therefore, the authors of [1] added the second stage introducing a weighting terms delivered by the first step described by (4):

$$\begin{aligned} \widehat{v}_t &= y_t^2 - \widehat{\varphi}_t^\top \widehat{\theta}_{t-1}, \\ \widehat{\theta}_t &= \widehat{\theta}_{t-1} + \frac{\widehat{P}_{t-1}\widehat{\varphi}_t \widehat{v}_t}{(\tilde{\varphi}_t^\top \tilde{\theta}_{t-1})^2 + \widehat{\varphi}_t^\top \widehat{P}_{t-1}\widehat{\varphi}_t}, \quad \widehat{\theta}_0 = 0, & (7)\\ \widehat{P}_t &= \widehat{P}_{t-1} - \frac{\widehat{P}_{t-1}\widehat{\varphi}_t \widehat{\varphi}_t^\top \widehat{P}_{t-1}}{(\tilde{\varphi}_t^\top \tilde{\theta}_{t-1})^2 + \widehat{\varphi}_t^\top \widehat{P}_{t-1}\widehat{\varphi}_t}, \quad \widehat{P}_0 = \kappa I, \end{aligned}$$

where $\widehat{\varphi}_t = \left(1, y_{t-1}^2, \widehat{\varphi}_{t-1}^\top \widehat{\theta}_{t-2}\right)^\top$, $\widehat{\theta}_t = (\omega, \alpha_1, \beta_1)^\top$, and again κ is a large number, e.g. $\kappa = 10^5$. The suggested time-varying weights should improve the efficiency of the estimates by reflecting the heteroskedastic structure of v_t, see [1]. Probably, this intention does not fulfil the expectation, compare with the conclusions in Section 4 and 5.

From the practical point of view, it is necessary to adopt a monitoring procedure that projects the recurrent estimates onto the region of stationarity. For instance, one can assume the technique recommended in [1] that completes the previous algorithms (see also [7] and [8]), i.e.:

1. Set $\mu \in [0, 1)$ and $i = 1$.
2. Check if $\tilde{\theta}_t$ (or $\widehat{\theta}_t$) fulfils the conditions of stationarity (see above). If yes, stop. Otherwise, go to 3.
3. Evaluate a new $\tilde{\theta}_t$ (or $\widehat{\theta}_t$): $\tilde{\theta}_t = \tilde{\theta}_{t-1} + \mu^i \frac{\tilde{P}_{t-1}\tilde{\varphi}_t \tilde{v}_t}{1+\tilde{\varphi}_t^\top \tilde{P}_{t-1}\tilde{\varphi}_t}$ $\left(\text{or } \widehat{\theta}_t = \widehat{\theta}_{t-1} + \mu^i \frac{\widehat{P}_{t-1}\widehat{\varphi}_t \widehat{v}_t}{(\tilde{\varphi}_t^\top \tilde{\theta}_{t-1})^2+\widehat{\varphi}_t^\top \widehat{P}_{t-1}\widehat{\varphi}_t}\right)$.
4. Return to 2 with $i \to (i + 1)$.

Although the introduced estimation schemes seem to be reliable at first glance, and the estimated parameters should converge to their true counterparts in some sense (according to [1], [7], and [9] using the corresponding assumptions), the results of numerical simulations presented in [1] are disputable. Therefore, we have decided to numerically revise the proposed algorithm in a Monte Carlo study, see Section 4. The obvious inaccuracies motivate further research, see Section 4 and 5.

4 Monte Carlo experiments

Various simulation experiments have been performed in the considered framework. In greater detail, several time series of different lengths with varying configurations of the parameters $(\omega, \alpha_1, \beta_1)$ were generated to analyse the considered algorithm assuming distinct values of the tuning parameters κ and μ. However, the obtained conclusions have remained almost analogical. Therefore, two representative settings are reviewed here. In particular, two Gaussian GARCH(1,1) processes were simulated 1000 times with the following specification of the parameters: (i) $(\omega, \alpha_1, \beta_1) = (0.10, 0.05, 0.90)$, (ii) $(\omega, \alpha_1, \beta_1) = (0.60, 0.20, 0.50)$. Obviously, the first set of the parameters determines the more persistent model than the second one. Additionally, different lengths of replicated time series were supposed, namely 100, 500, 1000, and 10000. All computations were conducted in `R` using $\kappa = 10^5$ and $\mu = 0.7$. It is noteworthy that the initial values of the estimated parameters appearing in (4) and (7) must have been slightly modified.

The recommended initialization vectors $(0, 0, 0)^\top$ do not lie in the region of stationarity. Accordingly, the original settings may lead to eventual errors. Therefore, both estimation algorithms were initialized with $(10^{-5}, 10^{-5}, 10^{-5})^\top$. The recursive estimates given by (4) and (7) have been compared with the common counterparts delivered by the conditional maximum likelihood estimation.

Table 1 summarizes the results of the considered Monte Carlo experiment for the first set of the true parameters, i.e. $(\omega, \alpha_1, \beta_1) = (0.10, 0.05, 0.90)$. In this case, the conditional maximum likelihood (ML) estimates demonstrate their expectable capabilities. With the increasing lengths of generated time series, the mean estimates are more precise and have decreasing standard deviations. On the other hand, the results of the introduced two stage recursive pseudo-linear regression (2SRPLR) estimation delivered by (4) and consequently by (7) are not indeed reliable. Firstly, the mean estimates do not converge to their true counterparts, compare with the maximum likelihood estimates. Secondly, the associated standard deviations, which express the consistency (and efficiency) of the estimates, are definitely not decreasing for $\hat{\omega}$ and $\hat{\beta}_1$. These findings contradict to the conclusions in [1]. Moreover, Figure 1 presents kernel densities of the calculated estimates for time series of the length 10000. Apparently, the recurrent estimates are strongly biased; compare with the maximum likelihood estimation case.

Case (i)	**100**	**500**	**1000**	**10000**
$\hat{\omega}^{2SRPLR}$	0.68095	0.72739	0.70855	0.78240
	(0.55770)	(0.56856)	(0.58342)	(0.61570)
$\hat{\alpha}_1^{2SRPLR}$	0.07851	0.04568	0.03820	0.02452
	(0.11420)	(0.06709)	(0.06167)	(0.03688)
$\hat{\beta}_1^{2SRPLR}$	0.10042	0.09360	0.09412	0.11498
	(0.17302)	(0.17950)	(0.19121)	(0.23721)
$\hat{\omega}^{ML}$	1.54239	0.84281	0.41899	0.10210
	(0.70845)	(0.79085)	(0.61690)	(0.02031)
$\hat{\alpha}_1^{ML}$	0.06021	0.05970	0.05590	0.04968
	(0.06645)	(0.04133)	(0.02848)	(0.00590)
$\hat{\beta}_1^{ML}$	0.16463	0.51283	0.73126	0.89924
	(0.26213)	(0.39715)	(0.31792)	(0.01427)

Table 1 Mean values of 2SRLPR and ML estimates and their standard deviations.

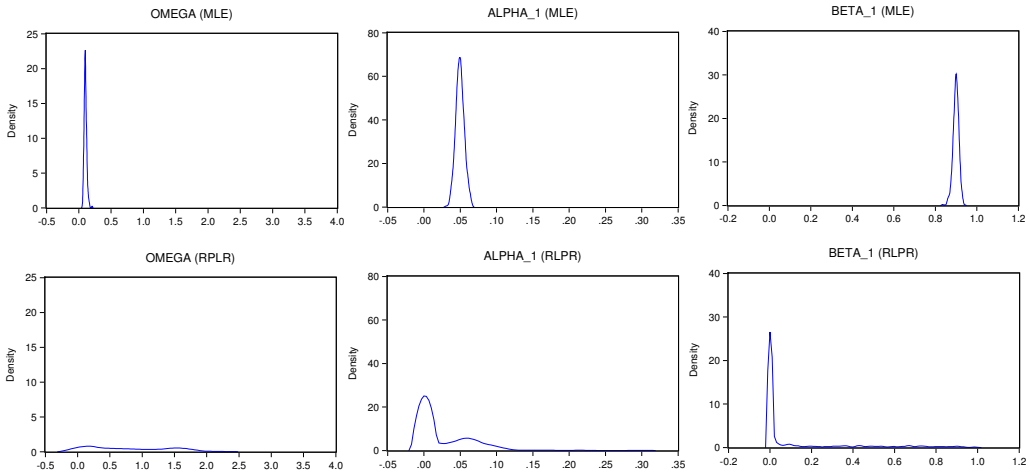


Figure 1 Kernel densities of 2SRPLR and ML estimates (the case (i), 10000 observations).

Analogously, one can view the results in Table 2 associated with the second set of the given parameters, i.e. $(\omega, \alpha_1, \beta_1) = (0.60, 0.20, 0.50)$. Again, it is clear that the studied recurrent algorithms do not deliver the acceptable mean estimates and that their standard deviations generally do not decrease. However, in this case, the conditional maximum likelihood estimation also fails comparing the mean estimates and

their true counterparts. It is well-known fact that in particular configurations of the GARCH(1,1) model parameters the maximum likelihood estimation technique can identify one of local extremes as the global solution of the objective function, see e.g. [6]. Nevertheless, this can be resolved e.g. by testing different initial values [1].

Table 3 and Table 4 include means of absolute values of $y_t^2 - \hat{\sigma}_{t,2SRPLR}^2$ and $y_t^2 - \hat{\sigma}_{t,ML}^2$ and their standard deviations for the first and second set of true parameters, respectively. These can be regarded as a specific measure of fit. From this point of view, both estimation techniques offer the comparable results; on the contrary, there are clear distinctions between the presented standard deviations, which indicate that the maximum likelihood procedure delivers more quality estimates.

Case (ii)	**100**	**500**	**1000**	**10000**
$\hat{\omega}^{2SRPLR}$	0.57205	0.57489	0.59273	0.64123
	(0.46461)	(0.43663)	(0.43952)	(0.44599)
$\hat{\alpha}_1^{2SRPLR}$	0.11227	0.09377	0.08899	0.07927
	(0.14666)	(0.11306)	(0.10816)	(0.09515)
$\hat{\beta}_1^{2SRPLR}$	0.09348	0.10534	0.12342	0.11771
	(0.15526)	(0.16914)	(0.19224)	(0.19443)
$\hat{\omega}^{ML}$	1.48350	1.41770	1.44367	1.74534
	(0.68208)	(0.60764)	(0.56289)	(0.14835)
$\hat{\alpha}_1^{ML}$	0.14830	0.16428	0.16362	0.16091
	(0.10478)	(0.04248)	(0.03138)	(0.00617)
$\hat{\beta}_1^{ML}$	0.14674	0.16817	0.14942	0.00737
	(0.25156)	(0.26433)	(0.24929)	(0.06203)

Table 2 Mean values of 2SRLPR and ML estimates and their standard deviations.

Case (i)	**100**	**500**	**1000**	**10000**
$y_t^2 - \hat{\sigma}_{t,2SRPLR}^2$	1.88327	1.86860	1.86830	1.87354
	(0.50554)	(0.25287)	(0.19471)	(0.20065)
$y_t^2 - \hat{\sigma}_{t,ML}^2$	1.91045	1.93572	1.93412	1.93544
	(0.54006)	(0.24367)	(0.18288)	(0.05699)

Table 3 Means of absolute values of $y_t^2 - \hat{\sigma}_{t,2SRPLR}^2$ and $y_t^2 - \hat{\sigma}_{t,ML}^2$ and their standard deviations.

Case (ii)	**100**	**500**	**1000**	**10000**
$y_t^2 - \hat{\sigma}_{t,2SRPLR}^2$	1.86566	1.87285	1.86265	1.86648
	(0.47421)	(0.24529)	(0.18257)	(0.10467)
$y_t^2 - \hat{\sigma}_{t,ML}^2$	1.95460	1.97949	1.96518	1.98799
	(0.57224)	(0.26420)	(0.18329)	(0.05975)

Table 4 Means of absolute values of $y_t^2 - \hat{\sigma}_{t,2SRPLR}^2$ and $y_t^2 - \hat{\sigma}_{t,ML}^2$ and their standard deviations.

5 Conclusions

In summary, the introduced results clearly show that the studied two stage recurrent estimation scheme recommended by [1] does not demonstrate the declared capabilities. On the other hand, the overall fit measures summarized by Table 3 and Table 4 suggest that the method can be revised to deliver better outcomes. This effort is essentially motivated by the necessity of on-line estimates, which can be crucial in many applications.

Consequently, the future research should be focused on the following issues. Firstly, one must reconsider weighting terms, compare (4) and (7), which are (from the author's point of view) the essential

elements influencing the whole technique. Generally, it is necessary to respect the heteroskedastic character of v_t in a more appropriate way. Secondly, one should take into account pragmatic approaches to on-line estimation presented by [7] and [9]. For instance, the supposed weighting scheme can include an exponentially decreasing term which adjusts the estimates and accelerates convergence, especially in the beginning of estimation. Thirdly, it may be possible to assume only a one stage algorithm to reduce computational complexity. Fourthly, various projections onto the region of stationarity should be compared. Lastly, a complex Monte Carlo study should be conducted to verify a suitable estimation procedure.

Acknowledgements

This research is supported by the grant GA P402/12/G097.

References

[1] Aknouche, A., and Guerbyenne, H.: Recursive estimation of GARCH models, *Communications in Statistics – Simulations and Computations*, **35** (2006), 925–938.

[2] Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31** (1986), 307–327.

[3] Bose, A., and Mukherjee, K.: Estimating the ARCH parameters by solving linear equations, *Journal of Time Series Analysis*, **24** (2003), 127–136.

[4] Durbin, J., and Koopman, S. J.: *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, 2012.

[5] Engle, R. F.: Autoregressive conditional heteroskedasticity and estimates of the variance of UK inflation, *Econometrica*, **50** (1982), 987–1008.

[6] Franke, J., Härdle, W. K. and Hafner, C. M.: *Statistics of Financial Markets: An Introduction*. Springer, Berlin, 2010.

[7] Ljung, L.: *System Identification: Theory for the User*. Prentice Hall, Upper Saddle River (NJ), 1998.

[8] Simon, D.: Kalman filtering with state constraints: A survey of linear and nonlinear algorithms, *Control Theory & Applications*, **8** (2010), 1303–1318.

[9] Söderström, T. S., and Stoica, P.G.: *System Identification*. Prentice Hall, Hertfordshire, 1989.

[10] Tsay, R. S.: *Analysis of Financial Time Series*. Wiley-Interscience, New York, 2005.

Saaty‘s matrix revisited: Securing the consistency of pairwise comparisons

Robert Hlavatý[1]

Abstract. The paper is intended to outline the ways of proper pairwise comparison matrix data filling. The decision makers are sometimes undisciplined while filling in the pairwise comparison matrix making their own pairwise comparison of criteria contradictory itself. The consistency of the matrix then reaches over a feasible limit and the information value of the data is ruined. The inconsistency originates in indiscipline or unawareness of the decision maker. The goal is to help or force the decision maker to fill the Saaty's matrix in a proper manner and make the data useful and consistent. For most common pairwise comparison matrices of the size 3x3 and 4x4, the intervals of stability are defined. If a Saaty's matrix appears to be inconsistent, the following text describes how the certain values in the matrix should be modified to receive a feasible result in terms of consistency.

Keywords: Saaty's matrix, decision making, inconsistency index, characteristic equation.

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

The tools for Multiple Criteria Decision Making (MCDM) have its use in various fields of human acting. According to Hwang and Yoon [3] The MCDM problem can be classified into two categories in this respect: Multiple Attribute Decision Making (MADM), and Multiple Objective Decision Making (MODM). In other words the set of decision alternatives can be either discrete finite or continuous infinite. In either of them, the criteria act as an important factor with essential impact on comparison of variants. Obviously, if all the criteria in a model had equal importance, the MCDM problem would be reduced to simpler problem. Usually, the importance of a criterion is expressed by its weight. The means of weights assessment were discussed many times. Entropy method proposed by Shannon [8] as well as the LINMAP method developed by Srinivasan and Shocker [10] can be applied if the decision matrix is known. These concepts are generally difficult to understand for a decision maker. When the pairwise comparison matrix is given, the Least square method and the Eigenvector method can be applied for weights assessment.

The easiest and most intuitive way of weights assessment in pairwise comparison matrix is use of various scoring scales. If one seeks to get "expert's opinion" in the form of weights then this is the most comfortable way of criteria evaluation for unqualified evaluators. It can be undertaken in a comparison matrix telling which one of two criteria is more important than the other and also to what extent is it preferred before the other. Saaty [5] ,[7] defined a pairwise comparison matrix in which, using the scale of $1/9$ to 9, the comparison of criteria can be done. The matrix itself must be reciprocal and the preference consistency must hold. Filling the preferences into the matrix is generally difficult to understand for common people whose opinions and preferences we are trying to obtain. It is even more difficult to force the decision makers to keep their opinions consistent, even for the low number (i.e. 3 or 4) of criteria compared.

After the preference information is gathered from the experts, the data might be damaged due to inconsistency or even missing because some of them had to be excluded due to inconsistency. Shirashi, Obata and Daigo [9] show the way of estimation of one missing datum in incomplete matrices using their own heuristic method. A different heuristic method for estimation of more missing data is described by Harker [2]. Also the fuzzy approach to preference in pairwise comparison matrix can be applied, as shown by Ramík [4]. It doesn't have to be necessary to exclude the damaged inconsistent data from the matrix. Inconsistency issues are widely discussed by Bozóki and Rapcsák [1] showing some interesting general properties of inconsistent matrices and their eigenvalues. In the following section, the analytical description of inconsistency is shown for 3x3 and 4x4 pairwise comparison matrices, using the brute-force approach to inference.

[1] University of Life Sciences, Department of Systems Engineering, Kamýcká 129, 165 21, Prague.

2 Material and methods

The idea is to find out how the certain values in pairwise comparison matrix can be changed while the consistency still holds. Saaty (1980) proposed a method for calculating inconsistency. The largest eigenvalue λ_{max} of pairwise comparison $n \times n$ matrix $\boldsymbol{A}$ is computed. It was shown [5] that $\lambda_{max} \geq n$ and is equal to n if and only if the matrix $\boldsymbol{A}$ is fully consistent. The **inconsistency index** was defined by

$$CI_n = \frac{\lambda_{max} - n}{n - 1} \tag{1}$$

The index alone has no true meaning until any critical values are introduced. At first, Saaty proposed that the inconsistency ratio lesser than 0,1 is generally acceptable. Later, Saaty [6] modified the critical value of inconsistency ratio to 0.05 for 3×3 matrices and 0.08 for 4×4 matrices. The values of 0.05 and 0.08 will be taken into account during the further calculations for corresponding matrices.

2.1 3×3 Pairwise comparison matrix

Let us have a general pairwise comparison matrix $\boldsymbol{A}$ of 3 criteria given by elements a, b, c:

$$\boldsymbol{A} = \begin{pmatrix} 1 & a & b \\ 1/a & 1 & c \\ 1/b & 1/c & 1 \end{pmatrix} \tag{2}$$

The characteristic polynomial of $\boldsymbol{A}$ is expressed as $det\ (\boldsymbol{A} - \lambda \boldsymbol{E})$=0. For the particular pairwise comparison matrix $\boldsymbol{A}$ we receive

$$\begin{vmatrix} 1-\lambda & a & b \\ 1/a & 1-\lambda & c \\ 1/b & 1/c & 1-\lambda \end{vmatrix} = 0 \tag{3}$$

and thus the characteristic equation of the matrix $\boldsymbol{A}$:

$$-\lambda^3 + 3\lambda^2 + \frac{b}{ac} + \frac{ac}{b} - 2 = 0 \tag{4}$$

It can be proved that the characteristic equation has the roots $\lambda_1 \in \mathbb{R} \wedge \lambda_1 \neq 0;\ \lambda_2, \lambda_3 \in \mathbb{C}$ or $\lambda_1, \lambda_2 \in \mathbb{R} \wedge \lambda_1 = 3, \lambda_2 = 0;\ \lambda_3 \in \mathbb{C}$ if the matrix $\boldsymbol{A}$ is fully consistent (i.e. $CI_n = 0$). Due to the nature of the cubic polynomial and the values of $a, b, c \in (0, 9\rangle$, it is possible to treat the λ as λ_{max}. There is an obvious dependency between the matrix elements a, b, c and the maximum eigenvalue λ. The matrix inconsistency depends on the entire triad of a, b, c values. Thus we can assume that for each combination $C(3,2)$ of a, b, c values there will be such third value that $CI_n = 0$. Assuming the inconsistency index must not reach over the threshold of CI_n, the formula (1) for this particular case must hold:

$$\frac{\lambda_{max} - 3}{3 - 1} \leq CI_n \tag{5}$$

$$\lambda_{max} \leq 2CI_n + 3 \tag{6}$$

Substituting the λ_{max} (6) for λ into the characteristic equation (4), it is possible to obtain graphical representation of dependency between either a, b, c and CI_n. Let us show at least one case e.g. $CI_n = f(a)$ with chosen $b = 6$ and $c = 6$. The function shows how the value of a affects the consistency index:

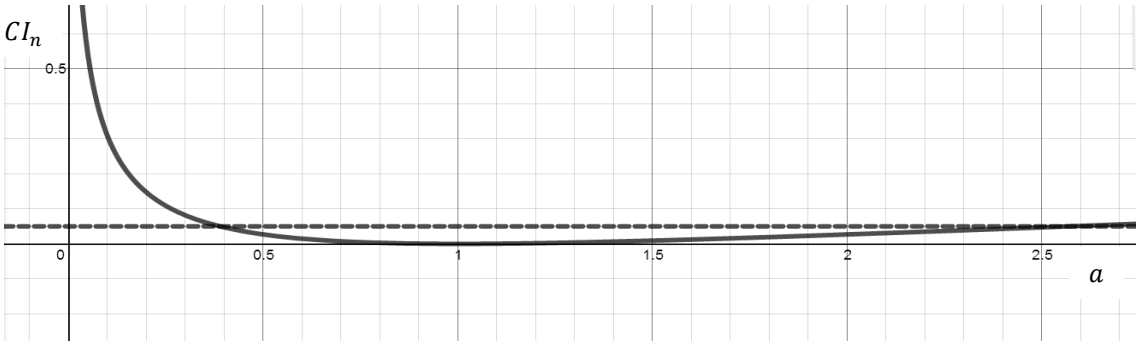


Figure 1 $CI_n = f(a), a \in \mathbb{R}^+$. (in Desmos graphing calculator)

The dash line expresses the threshold value of 0.05 for inconsistency index. If the function reaches under the threshold then $CI_n \leq 0.05$. The intercepts of the function with the threshold show the range of a as the consistency holds. The range for this particular case is $a \cong \langle 0.39\,, 2.57 \rangle$. The lower and upper bound of a express the critical value of a where the $CI_n = 0.05$. The local minimum of $f(a)$ lies in [1,0] expressing the ideal function of the a value where the $CI_n = 0$. Practically, going back to the pairwise comparison matrix, the value of a should lie in the range $\langle 0.39\,, 2.57 \rangle$. Using the Saaty's scale, the a should be then chosen from the following set of numbers: $a \in \left\{\frac{1}{2}, 1, 2\right\}$ so the inconsistency of the matrix $\boldsymbol{A}$ would be acceptable. The same principle works for expressing $CI_n = f(b)$ and $CI_n = f(c)$. Clearly, it is possible to define such feasible interval for either a, b or c if the two others from the triad are known and the acceptable inconsistency index CI_n is set. Also the ideal value for zero inconsistency can be calculated.

For practical application it is desirable to calculate lower and upper bounds for either a, b and c. It is possible to express a, b, c from the characteristic equation (4). Since the parameters a, b, c in the equation are found in the fractions' numerators and denominators at the same time, the expression of these parameters will lead to quadratic equations. The solutions of these equations are expressed as follows:

$$\begin{aligned} a &= \frac{b(\lambda^3 - 3\lambda^2 + 2 \pm \sqrt{\lambda^6 - 6\lambda^5 + 9\lambda^4 + 4\lambda^3 - 12\lambda^2}}{2c} \\ b &= \frac{ac(\lambda^3 - 3\lambda^2 + 2 \pm \sqrt{\lambda^6 - 6\lambda^5 + 9\lambda^4 + 4\lambda^3 - 12\lambda^2}}{2} \\ c &= \frac{b(\lambda^3 - 3\lambda^2 + 2 \pm \sqrt{\lambda^6 - 6\lambda^5 + 9\lambda^4 + 4\lambda^3 - 12\lambda^2}}{2a} \end{aligned} \tag{7}$$

The polynomial in discriminant returns positive values for all $\lambda > 3$. This must hold true since it was proved that $\lambda_{max} \geq n$.Considering the polynomial in discriminant it can be said that either of these equations will always have 2 different solutions in $\mathbb{R}$ or one solution in $\mathbb{R}$ of multiplicity 2 if $= 3$, i.e. the inconsistency index is 0. The range of feasible values of a, b, c will always lie within the range of two roots of each quadratic equation.

2.2 4×4 pairwise comparison matrix

Let us describe the situation for matrices of the size 4×4. A general pairwise comparison matrix $\boldsymbol{B}$ given by elements a, b, c, d, e, f is constructed:

$$\boldsymbol{B} = \begin{pmatrix} 1 & a & b & c \\ 1/a & 1 & d & e \\ 1/b & 1/d & 1 & f \\ 1/c & 1/e & 1/f & 1 \end{pmatrix} \tag{8}$$

The characteristic polynomial of $\boldsymbol{B}$ is expressed as $det\ (\boldsymbol{B} - \lambda \boldsymbol{E})$=0. For the particular pairwise comparison matrix $\boldsymbol{B}$ we receive

$$\boldsymbol{B} = \begin{pmatrix} 1-\lambda & a & b & c \\ 1/a & 1-\lambda & d & e \\ 1/b & 1/d & 1-\lambda & f \\ 1/c & 1/e & 1/f & 1-\lambda \end{pmatrix} \tag{9}$$

The expression of the characteristic equation becomes somehow difficult and time consuming, although achievable. The Laplace expansion $(-1)^{i+j}M_{ij}$ for ij-th minor of $\boldsymbol{B}$ is used and extensive characteristic equation is found:

$$\lambda^4 - 4\lambda^3 + (8 - C')\lambda + C'' = 0 \tag{10}$$

where

$$C' = \frac{e}{df} + \frac{df}{e} + \frac{c}{ae} + \frac{ae}{c} + \frac{b}{ad} + \frac{ad}{b} + \frac{c}{bf} + \frac{bf}{c} \tag{11}$$

and

$$C'' = C' - \frac{adf}{c} - \frac{c}{adf} - \frac{ae}{bf} - \frac{bf}{ae} - \frac{be}{cd} - \frac{cd}{be} - 2 \tag{12}$$

It can be proved that the characteristic equation (10) has the roots $\lambda_1, \lambda_2 \in \mathbb{R} \wedge \lambda_1 \neq 0$; $\lambda_3, \lambda_4 \in \mathbb{C}$ or $\lambda_1, \lambda_2 \in \mathbb{R} \wedge \lambda_1 = 4, \lambda_2 = 0$; $\lambda_3, \lambda_4 \in \mathbb{C}$ if the matrix $\boldsymbol{B}$ is fully consistent (i.e. $CI_n = 0$). Again, the λ from the characteristic equation can be treated as λ_{max}. This time, however, the assumption is not true that for every five values of a, b, c, d, e, f there would be the sixth value such that $CI_n = 0$. As the polynomial became more complex than in the case of 3×3 matrix, it may happen that for five known values of a, b, c, d, e, f there is no such remaining value that would make the whole matrix fully consistent. The dependency between a random value of a, b, c, d, e, f and the consistency index CI_n can be shown as graphical representation. Using the inconsistency formula (1) we can express λ_{max} as

$$\lambda_{max} \leq 3CI_n + 4 \tag{13}$$

and then substitute λ_{max} (13) for λ into the characteristic equation (10). Let us show two examples, the first e.g. $CI_n = f(a)$ with chosen $b = 3, c = 5, d = 5, e = 6, f = 3$. Then the relation shows how the value of a affects consistency index:

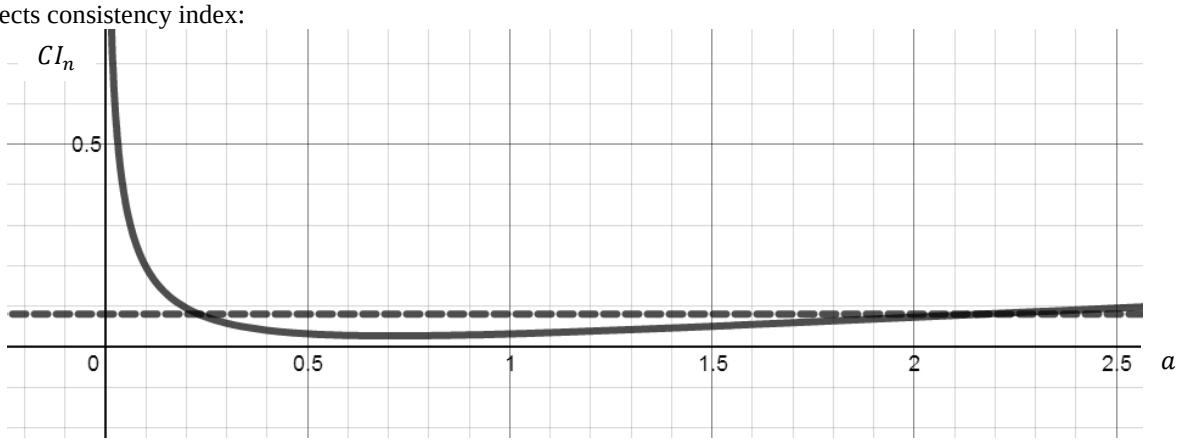


Figure 2 $CI_n = f(a), a \in \mathbb{R}^+$. (in Desmos graphing calculator)

The dash line again expresses the threshold value of 0.08. The intercepts of the dash line with the curve express the range of a as the consistency holds. The actual range is $a \cong \langle 0.23\ ,\ 2.17 \rangle$. Compared to the 3×3 case, the curve will not reach the value of zero and so there is no such value of a making the matrix fully consistent. Although the local minimum of the curve represents the value of a at which the CI_n reaches its minimum for positive values of a. We will now change the parameter $d = 5$ to $d = 1/9$. The situation changes as depicted on the following picture:

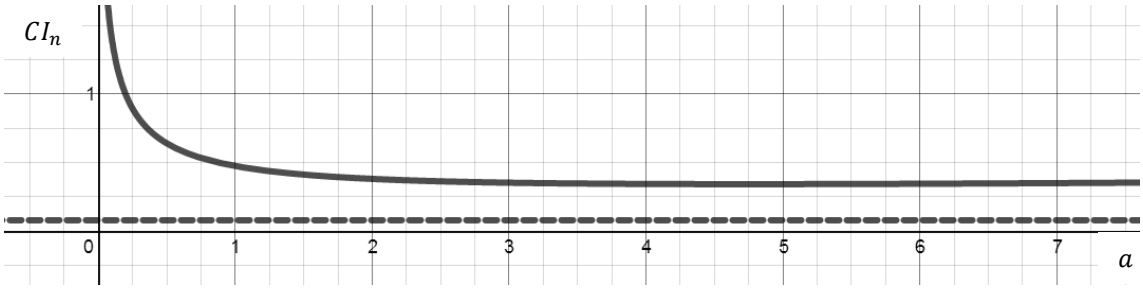


Figure 3 $CI_n = f(a), a \in \mathbb{R}^+$. (in Desmos graphing calculator)

Clearly the curve does not cross the threshold of 0,08, at least not in terms of positive numbers. Theoretical values of a for this case could be found in the range $a \cong \langle -6.97, -3.22 \rangle$ where the $CI_n \leq 0.08$ is satisfied. This is not applicable since we operate only with positive values of preference. Practically, there is no such value of a that would satisfy the inconsistency index and more values of a, b, c, d, e, f had to be changed at once.

Once again, it is desirable to determine the feasible range for the values of a, b, c, d, e, f so the inconsistency index would be acceptable. Expressing individual parameters a, b, c, d, e, f from the characteristic equation (10) results in complicated expressions that complete a quadratic equation. For instance the expression of parameter a is shown:

$$a = \frac{-l \pm \sqrt{l^2 - 4km}}{2k} \tag{14}$$

where

$$\begin{gathered} k = \lambda bde^2 f + \lambda cd^2 ef - bd^2 ef^2 - cde^2 \\ l = (\lambda^4 - 4\lambda^3 + 8\lambda - 2)bcdef + \lambda bce^2 + \lambda c^2 de + \lambda b^2 def^2 + \lambda bcd^2 f^2 - b^2 e^2 f - c^2 d^2 f \\ m = \lambda b^2 cef + \lambda bc^2 df - b^2 cdf^2 - bc^2 e \end{gathered} \tag{15}$$

The expressions of b, c, d, e, f will be not shown because they lead to very similar complicated results, only with differing parameters of a, b, c, d, e, f.

3 Results and discussion

It was shown how the values in the pairwise comparison matrix influence the inconsistency index. In case of 3×3 matrix it is always possible to define the triad of parameters a, b, c such that the inconsistency index is zero. Although in some cases the triad of a, b, c values cannot be practically used for weights estimation. For example we can consider the following pairwise comparison matrix of three criteria with known parameters b and c:

$$\boldsymbol{A} = \begin{pmatrix} 1 & a & 9 \\ 1/a & 1 & 1/9 \\ 1/9 & 9 & 1 \end{pmatrix} \tag{16}$$

Obviously if the preference of 1st criterion to 3rd criterion is very strong (i.e. =9) and simultaneously, the preference of the 3rd criterion to the 2nd is very strong, then the preference of the 1st criterion to the 2nd would have to be "very strong × very strong" (i.e. $9 \times 9 = 81$). We shall use the formula (7) for estimation of the value a so the matrix inconsistency would be acceptable. We receive the range of $a \cong \langle 21.4\,, 306.5 \rangle$ and the ideal value of $a = 81$ as expected. This is practically unacceptable since the scale of $\frac{1}{9} \ldots 9$ is used. This kind of decision maker's preference is wrong in its own nature as the 2nd criterion here has negligible importance and perhaps should not be included into decision making process at all.

In case of 4×4 matrix, the way of preference filling into matrix appears to be more complicated when an acceptable inconsistency index is desired. It emerged from the properties of the characteristic equation (10) and formula (14) that 4 situations could possibly occur for estimation of the parameters a, b, c, d, e, f. Let us denote an arbitrary parameter of a, b, c, d, e, f as the general parameter p with the ideal value p^* for $CI_n = 0$ and acceptable range $p \in \langle \underline{p}\,,\ \overline{p} \rangle$ for $CI_n \leq 0.08$. The situations are then following:

- $p^* \in \langle \underline{p}\,,\ \overline{p}\rangle \cap \langle \frac{1}{9}, 9\rangle \wedge p \in \langle \underline{p}\,,\ \overline{p}\rangle \cap \langle \frac{1}{9}, 9\rangle$. If the value of p is set within the range, than the inconsistency value is acceptable. Simultaneously the ideal value of p^* is found where the inconsistency index $CI_n = 0$. This situation would appear rarely as the ideal value usually cannot be found due to complexity of the characteristic equation. Full consistency of the matrix can be reached by changing of the single parameter.
- $p^* \notin \langle \underline{p}\,,\ \overline{p}\rangle \cap \langle \frac{1}{9}, 9\rangle \wedge p \in \langle \underline{p}\,,\ \overline{p}\rangle \cap \langle \frac{1}{9}, 9\rangle$. If the value of p is set within the range, than the inconsistency value is acceptable. Ideal value of p^* is not found. It is not possible to reach full consistency of the matrix only by changing a single parameter.
- $p^* \notin \langle \underline{p}\,,\ \overline{p}\rangle \cap \langle \frac{1}{9}, 9\rangle \wedge p \in \langle \underline{p}\,,\ \overline{p}\rangle \cap (-\infty\,,\ \frac{1}{9}) \cup (9\,, \infty)$. The consistency is acceptable only for such parameter value p out of the feasible range.
- $p^* \notin \mathbb{R} \wedge p \notin \mathbb{R}$. The range of the parameter p cannot be found because the solution of (10) has no real roots. There is no such value of p that would make the consistency of the matrix acceptable.

The four situations show expressions of but one parameter of the sextuplet a, b, c, d, e, f. If all the parameters are expressed at the same time, the six intervals will be shown for each of them. It is than necessary to find such intervals where the 1st or 2nd situation of above occurred. Changing the appropriate parameter will make the inconsistency index acceptable.

4 Conclusion

The way how to treat preference values in the 3×3 and 4×4 pairwise comparison matrices was proposed. The small size matrices were subjected to analysis because the higher the size is the more difficult it gets to hold on to consistent opinions. The methodology could be as well applied for matrices of larger size with some difficulties. Finding the characteristic equation for larger matrix can be conceivable using the Newton's identities, although this task becomes very time consuming as the matrix size grows larger. Even for the matrix of 4×4 size the expression of the individual parameters seems to be complicated but if the formulae are implemented into the computer environment (e.g. MS Excel will do nicely) than we gain an easy tool for consistency evaluation. If a decision maker's preferences are gathered and placed into pairwise comparison matrix, the inconsistency check should be done afterwards. Should the inconsistency index of the matrix be unacceptable then the feasible intervals for single preference values can be found. These intervals will show the problematic preference values and in addition, how these values could be possibly changed ex-post – if it is possible for a decision maker to do so.

References

[1] Bozóki, S., and Rapcsák, T.: On *Saaty's* and *Koczkodaj's* inconsistencies of pairwise comparison matrix. *Journal of Global Optimization* **42** (2008), 157–175.

[2] Harker, P. T.: Alternative modes of questioning in analytic hierarchic process. *Mathematical modelling* **9** (1987), 353–360.

[3] Hwang, C.L. and Yoon, K.: *Multiple attribute decision making:methods and applications: a state-of-the-art survey*. Springer-Verlag, 1981.

[4] Ramík, J.: Fuzzy preference matrix with missing elements and its application to ranking of alternatives. *Mathematical Methods in Economics* **31**(2013), 767-772.

[5] Saaty, T.L.: *The Analytic Hierarchy Process*. McGraw-Hill, 1980.

[6] Saaty, T.L.: *Fundamentals of decision making*. RSW Publications, 1994.

[7] Saaty, T.L: Relative measurement and its generalization in decision making. *RACSAM*, **102(2)** (2008), 251–318.

[8] Shannon, C. E.: Prediction and entropy of printed English. *Bell system technical journal* **30** (1951), 50–64.

[9] Shiraishi, S., Obata, T.,Daigo, M.: Properties of a positive reciprocal matrix and their application to AHP. *Journal of the operations research* **41** (1998),404–414.

[10] Srinivasan, V., and Shocker A.D.: Linear programming techniques for multidimensional analysis of preference. *Psychometrika* **38** (1973), 337–342.

Zero lower bound on interest rate: application of DSGE model on Czech economy

Miroslav Hloušek[1]

Abstract. This paper deals with impacts of zero lower bound of interest rate on behaviour of macroeconomic variables in the Czech economy. The analysis uses DSGE model of small open economy, concretely model from Justiniano and Preston [4], and toolkit from Guerrieri and Iacoviello [2] that is able to solve occasionally binding constraint. The model is estimated on data of the Czech economy covering period 1996:Q2 – 2013:Q4. The behaviour of the model is studied using impulse responses and simulations in reaction to model specific shocks. Situation when central bank is incapable to decrease interest rate because of zero lower bound has implications for behaviour of output and consumption but not inflation. Shocks that cause binding of interest rate for most periods are domestic cost-push shock and foreign preference shock. These two shocks also have non-negligible welfare implications regarding behaviour of consumption. In reaction to three standard deviations of innovations, the drop of consumption was larger by 0.9 and 0.5 percentage points, respectively, when the interest rate hit the zero lower bound.

Keywords: zero lower bound on interest rate, DSGE model, occasionally binding constraint

JEL classification: E37
AMS classification: 91B51

1 Introduction

Monetary policy in many countries, including the Czech Republic, has been recently confronted with similar problem. Economy is in recession or slowly recovers from it, inflation is falling, but nominal interest rates are close to zero and could not be further lowered to boost the economy. Nominal interest rate is limited by zero value and could not be negative. Naturally, this fact has consequences for the whole economy and it was also motivation for this paper. It explores situation in the Czech economy for case of zero lower bound of interest rate. Concretely, I use DSGE model estimated on the Czech data and then run series of simulations. I investigate for which shocks, what size of shocks and for how many periods the zero lower bound matters and what are implications for inflation and consumption. The results show that the most distinct shocks are domestic cost-push shock and foreign preference shock. The interest rate was binding in most cases and differences in consumption between binding and not binding case reached 0.9 and 0.5 percentage points for these two shocks. On the other hand, there were no differences for development of inflation between simulations for any shock.

The rest of the paper is organized as follows. Section 2 briefly describes the structure of the model, Section 3 comments on data and estimation and simulation techniques. Results of the estimation are described in Section 4 and Section 5 discusses the main findings of the analysis. Final section concludes with prospects for further research.

2 Model

The model is borrowed from Justiniano and Preston [4]. It is model of small open economy. The log-linearized model consists of fourteen structural equations and definitions for domestic and foreign economy and five laws of motion for exogenous shocks. It is New Keynesian model with nominal and real rigidities. There is habit in consumption of households and monopolistic price setting on market of goods which results in two Phillips curves (for domestic and imported goods). The monetary policy follows Taylor rule with interest rate smoothing.

[1]Masaryk University, Department of Economics, Lipová 41a, 602 00 Brno, hlousek@econ.muni.cz

The foreign block is modelled analogously, but is little bit simpler. The model contains seven shocks: cost-push, preferences and monetary shocks for both economies and risk premium shock. The shocks follow AR(1) process except of monetary shocks which are *iid*. Log-linearized model equations are quoted in the web Appendix [3].

3 Data and techniques

The model is estimated using data for variables of domestic economy – output, inflation and nominal interest rate, their foreign counterparts, and the real exchange rate. The model is expressed in gap form, therefore the time series are detrended using Hodrick-Prescott filter. Only the inflation rate (the growth rate of prices) is demeaned. I use quarterly data from the Czech Statistical Office, the Czech National Bank and the Eurostat databases which cover time period 1996:Q2 – 2013:Q4. A detailed description can be found in the web Appendix [3].

The model parameters are estimated using Bayesian techniques. Posterior distribution of the parameters was obtained by Random Walk Chain Metropolis-Hastings algorithm. It generated 300,000 draws in two chains with 150,000 replications each, 60 % of replications were discarded so as to avoid influence of initial conditions. MCMC diagnostics were used for verification of the algorithm. All computations were carried out using Dynare toolbox (Adjemian et al. [1]) in Matlab software.

Subsequent simulations are carried out by *Occbin* toolbox developed by Guerrieri and Iacoviello [2]. The authors use piecewise linear perturbation solution for dynamic models with occasionally binding constraint. They show that this method provides very good approximation of dynamic programming solution but it is much easier for implementation. It can capture nonlinearities that arises in models with two regimes – binding and not-binding constraint. Therefore this method allows to study the effects of attaining zero lower bound on nominal interest rate in broad range of models.

4 Results of estimation

The only parameter that is not estimated is discount factor β. It is calibrated to 0.9963 according to the mean of the real interest rate (1.49% annualized) calculated from data. The prior and posterior distribution of estimated parameters are quoted in Table 1. The results are quite common and correspond to other studies for the Czech economy (see e.g. Slanicay [5]). Calvo parameters for domestic and imported goods prices are virtually the same and are not far from the priors, $\theta_h = \theta_f = 0.71$. However, indexation of imported goods is much higher ($\delta_f = 0.44$, $\delta_h = 0.08$). Calvo parameter of prices in Euroarea is higher than in the Czech Republic, $\theta^* = 0.86$, but the indexation is low, $\delta^* = 0.06$. Parameter expressing the degree of openness, $\alpha = 0.63$, indicates that the share of imported goods in domestic consumption basket is 63 %. Habit formation parameter, $h = 0.73$, is little bit lower than the prior mean but still in accordance with empirical studies for the Czech economy. Parameter $\varphi = 3.12$ indicates quite low elasticity of labour supply. Similarly, elasticity of substitution between domestic and foreign goods is low, $\eta = 0.13$. Parameters in monetary policy rules show that the Czech central bank pays higher attention to output gap than ECB ($\psi_y = 0.56$ vs. $\psi_y^* = 0.26$). Weight to inflation is the same, $\psi_\pi = \psi_\pi^* = 1.56$, but interest rate smoothing parameter is slightly higher in Euroarea ($\rho = 0.81$ vs. $\rho^* = 0.89$). From the autoregressive parameters of shocks it is worth to mention very low persistence of risk premium shock ($\rho_s = 0.29$). Other shocks are much more persistent; AR parameters are around 0.7. The standard deviations of the shocks are quoted in percent in the lower part of the table. The most volatile shock is domestic cost-push shock, $\sigma_{e_a} = 6.99$, followed by foreign cost-push shock, $\sigma_{e_a^*} = 6.27$. On the other hand, the least volatile shocks are foreign and domestic monetary shocks ($\sigma_{e_m} = 0.14$ and $\sigma_{e_m^*} = 0.67$).

5 Simulations

The shocks and their impacts for model behaviour are very important in our analysis. First, I will use impulse response functions to study reaction of the variables in responses to positive and negative values of the shocks. Second, I will use smoothed innovations acquired from estimation and simulate the model behaviour in reaction to these series of shocks.[2] Finally, I use random number generator to generate series of (one thousand) innovations multiplied by corresponding standard deviation and again simulate behaviour of the model variables. The subject of our interest is, if the nominal interest rate hits the zero lower bound, for how many periods it is binding and if

[2]Most shocks follow AR(1) process, e.g. $\varepsilon_{g,t} = \rho_g \varepsilon_{g,t-1} + e_{g,t}$. I use term *shock* for the variable $\varepsilon_{g,t}$ and *innovation* for $e_{g,t}$ which has properties $e_{g,t} \sim N(0, \sigma_{e_g})$.

		Prior distribution		Posterior distribution		
Param.	Interpretation	Mean	S.D.	Mean	2.5 %	97.5 %
Price rigidities						
θ_h	Calvo, domestic goods	0.7	0.05	0.71	0.67	0.76
θ_f	Calvo, imported goods	0.7	0.05	0.71	0.64	0.78
θ^*	Calvo, foreign goods	0.8	0.05	0.86	0.82	0.89
δ_h	indexation, dometic goods	0.5	0.20	0.08	0.01	0.14
δ_f	indexation, imported goods	0.5	0.20	0.44	0.10	0.76
δ^*	indexation, foreign goods	0.5	0.20	0.06	0.01	0.11
Preference parameters						
α	openness	0.7	0.05	0.63	0.56	0.70
h	habit formation	0.8	0.05	0.73	0.67	0.79
σ	inverse elast. of intertemp. subst.	1.5	0.20	1.36	1.07	1.64
φ	inverse elast. of labor supply	2.5	0.50	3.12	2.41	3.85
η	elasticity of substitution between domestic and foreign goods	0.3	0.10	0.13	0.08	0.19
Monetary policy rules						
ψ_π	weight to inflation	1.5	0.15	1.56	1.31	1.82
ψ_y	weight to output gap	0.5	0.15	0.60	0.40	0.79
ρ	interest rate smoothing	0.8	0.05	0.82	0.78	0.86
ψ_π^*	foreign, weight to inflation	1.5	0.15	1.56	1.31	1.81
ψ_y^*	foreign, weight to output gap	0.5	0.15	0.27	0.14	0.38
ρ^*	foreign, interest rate smoothing	0.8	0.05	0.89	0.87	0.92
Persistence of of shocks						
ρ_g	AR param., shock to preferences	0.8	0.05	0.74	0.66	0.82
ρ_a	AR param., cost-push	0.8	0.05	0.72	0.63	0.81
ρ_s	AR param., risk premium	0.5	0.20	0.29	0.14	0.45
ρ_g^*	AR param., f. preferences	0.8	0.05	0.73	0.66	0.81
ρ_a^*	AR param., f. technology	0.8	0.05	0.69	0.60	0.79
Volatility of shocks						
σ_{e_g}	preferences	1.0	Inf	2.15	1.50	2.80
σ_{e_a}	cost-push	1.0	Inf	6.99	4.73	9.53
σ_{e_s}	risk premium	1.0	Inf	2.41	1.73	3.08
σ_{e_m}	monetary	1.0	Inf	0.67	0.56	0.78
$\sigma_{e_a^*}$	f. cost-push	1.0	Inf	6.27	3.56	9.18
$\sigma_{e_m^*}$	f. monetary	1.0	Inf	0.14	0.12	0.16
$\sigma_{e_g^*}$	f. preferences	1.0	Inf	0.67	0.56	0.78

Table 1 Results of estimation

it has implications for other model variables. Results of these experiments are summarized in Table 2. However, before going to the statistics in the table we will examine reactions in the graphs.

Figure 1 shows reaction of the model variables to domestic cost-push shock. The size of the shock is two standard deviations of innovation (posterior mean obtained from estimation). The y-axis shows percentage deviation from the steady state except of the first subgraph which is in percent level. The dashed red line depicts linear solution and corresponds to situation when nominal interest rate can be also negative, the solid blue line depicts piecewise linear solution and corresponds to zero lower bound case. Let us examine the linear solution first. The economy is subject to negative cost-push shock in period five. It causes decrease of inflation and central bank reacts by lowering interest rate. After small initial drop, the consumption rises above its steady state and output rises as well because of low interest rate. In piecewise linear solution, the central bank cannot decrease nominal interest rate enough because it hits the zero value. It is interesting that it does not have impact for inflation; its reaction is the same in both cases. However, reaction of consumption and output is different. Consumption falls much more and increase of output is smaller. After twenty periods the variables are back on their steady state values. Then, in period twenty five, positive cost-push shock hits the economy. The reaction of the variables is mirror image of the

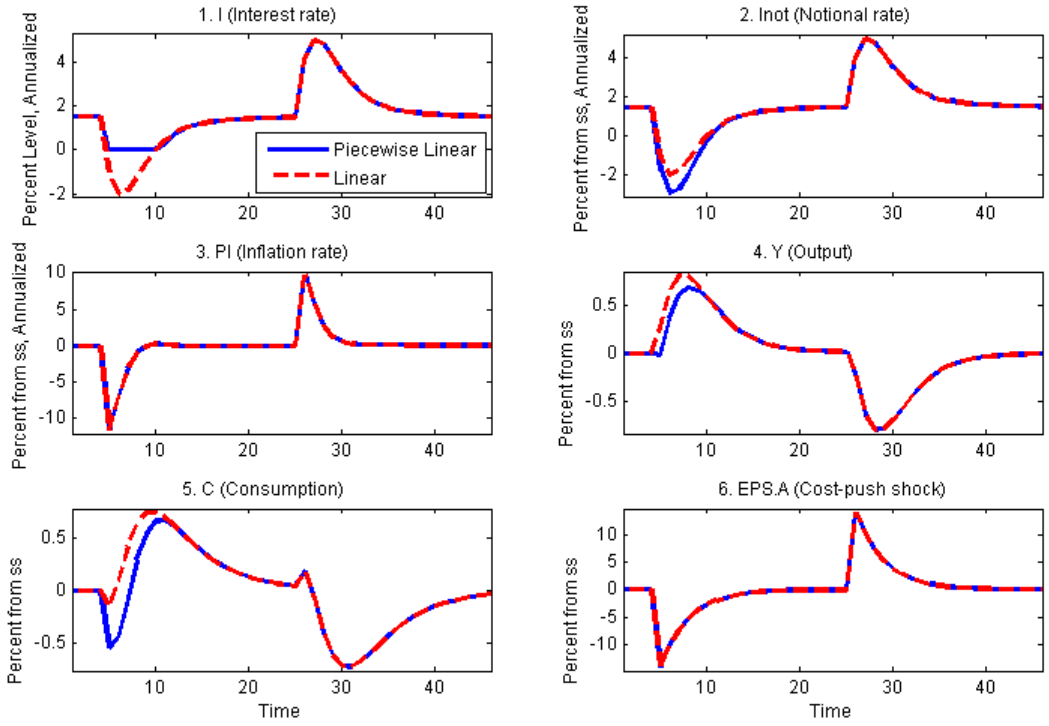


Figure 1 Impulse responses, cost-push shock, $2\sigma_{e_a}$

previous case, but now both solutions coincide.

Columns four to six in Table 2 show in how many periods the interest rate hits zero lower bound during the impulse response. It is calculated for three values of shocks: one, two and three standard deviations of innovations. It is interesting that foreign cost-push shock has quite large standard deviation, but the constraint is not binding in any case. There is also no binding of constraint for foreign monetary shock of any size. In reaction to average value of the shocks (one standard deviation of innovation) only domestic cost-push shock and monetary policy shock cause binding interest rate. Going to higher values of innovations, domestic cost-push shock plays important role. Interest rate hits the zero bound for six periods in reaction to two standard deviations of innovations.[3] Also foreign preference shock is becoming important for higher values of the innovations. On the other hand, domestic monetary policy shock cause binding of interest rate for one or two periods without much connection to the size of the innovation.

Shock	Interpretation	Posterior mean (in %)	Constraint binds irf (periods) $1\sigma_e$	$2\sigma_e$	$3\sigma_e$	smoothed shocks (%)	simulated shocks (%)	impact on C ZLB-noZLB at $3\sigma_e$ (p.p.)
σ_{e_g}	preferences	2.51	0	0	4	3	14	0.08
σ_{e_a}	cost-push	6.99	2	6	7	15	34	0.91
σ_{e_s}	risk premium	2.41	0	0	2	10	7	0.03
σ_{e_m}	monetary	0.67	1	2	2	15	23	0.21
$\sigma_{e_a^*}$	f. cost-push	6.27	0	0	0	0	8	0.00
$\sigma_{e_m^*}$	f. monetary	0.14	0	0	0	0	0	0.00
$\sigma_{e_g^*}$	f. preferences	0.67	0	4	8	28	27	0.52

Table 2 Binding of constraint in reaction to shocks

Figure 2 shows reactions of the variables for historically estimated innovations of cost-push shock. Again we can see the difference between binding and not binding constraint of interest rate. And similarly to impulse

[3]This reaction was depicted in Figure 1.

responses, there is no difference for behaviour of inflation, but trajectories of consumption and output differ in some cases.

Seventh column of Table 2 counts in how many percent the zero lower bound of interest rate is reached. From this historical perspective, the most important shocks were foreign preference shock followed by domestic cost-push and monetary policy shocks. The eight column shows the same measure as the previous one, but for simulated innovations (one thousand periods). The pattern is similar, but the ordering of the shocks is little different. Now the most important shock is domestic cost-push shock followed by foreign preference shock and domestic monetary shock. However, this simulation highly depends on sequence of random draws and must be taken with caution.

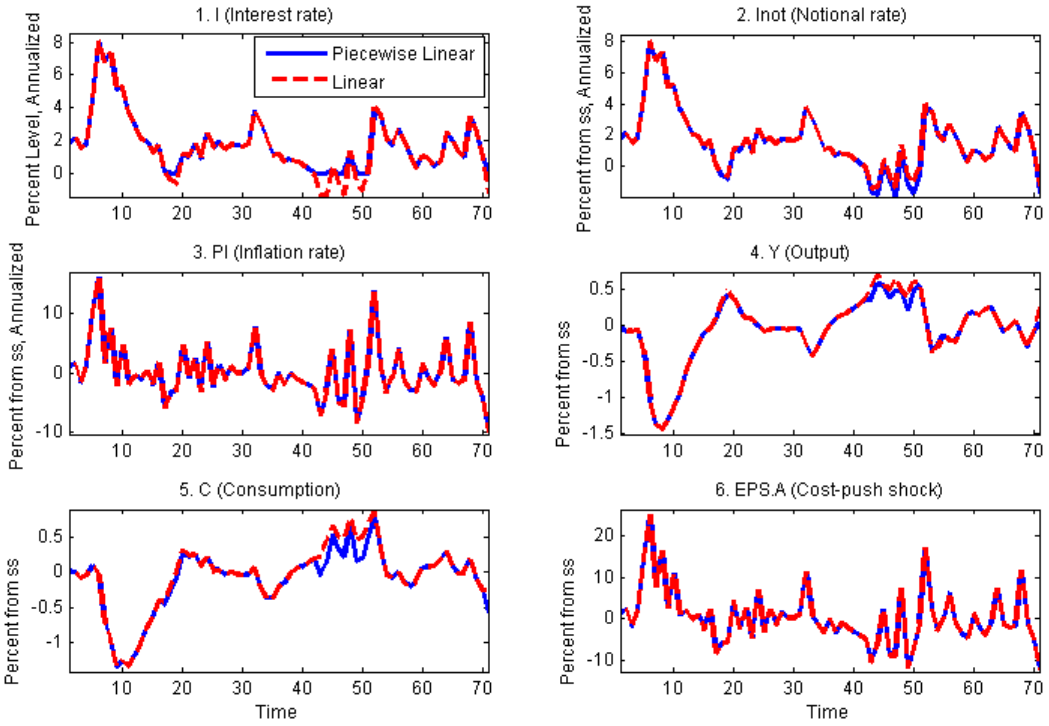


Figure 2 Simulation of variables in reaction to smoothed cost-push shock

Finally, we will quantify impacts of zero lower bound for behaviour of consumption. Figure 3 shows reaction of consumption to cost-push shock for different values of innovations (up to three standard deviations, both positive and negative values). The upper panel shows impact for consumption at peak (trough) of impulse responses in reaction to the shock for binding and not binding interest rate. The size of the shock in percent is on x-axis. The y-axis measures the deviation of consumption from steady state (also in percent). The lower panel depicts difference between two lines in the upper panel.

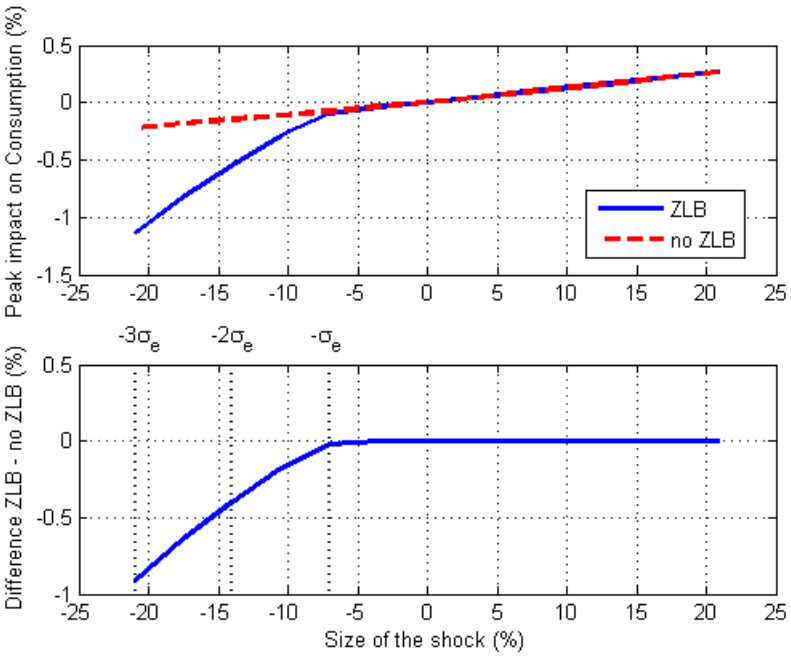
We see that the difference is increasing nonlinearly with the size of the shock. At three standard deviations of innovation, the drop in consumption in zero lower bound case is bigger by 0.91 percentage points against the case without binding interest rate.[4] The differences for other shocks are quoted in last column of Table 2. We see that foreign preference shock and domestic monetary shock have quite non-negligible impact for consumption – drop is bigger by 0.52 and 0.21 percentage points, respectively. However, the situation that the economy is hit by shocks of such magnitude is not so likely. Given the fact that the innovations have normal distribution, the shocks higher than three standard deviations are probable in 0.3 % of cases.

6 Conclusion

This paper explored possibility of attaining zero lower bound of interest rates in the Czech economy through estimated DSGE model. Several model specific shocks turned out as important for behaviour of nominal interest

[4]Difference at two standard deviations is 0.4 p.p.

Figure 3 Response of consumption to cost-push shock (up to $3\sigma_{e_a}$)



rate. The most significant seems domestic cost-push shock, monetary policy shock and foreign preference shock. Zero lower bound on interest rate has no implication for behaviour of inflation but it influences consumption and output. If the central bank cannot decrease interest rate because of zero lower bound, consumption falls more by 0.9 percentage points for high values of cost-push shock. Little bit lower consumption losses were found for foreign preference shock and domestic monetary shock. However, the probability that such big shocks hit the economy is quite low. Recent experience in the Czech economy also showed the central bank can use alternative instrument – exchange rate interventions. On the other hand, the analysis omits situations that the economy is subject to series of different shocks which can have multiplicative effect. However, given the computational possibilites, these impacts are hard to quantify.

To summarize it, we can say that the threat of attaining zero lower bound is quite substantial, but the welfare consequences are only moderate. Obviously, the results may depend on the model used. Therefore further research will examine impacts of zero lower bound in different model structure.

Acknowledgements

This paper is supported by specific research project no. MUNI/A/0775/2013 at Masaryk University.

References

[1] Adjemian, S., Bastani, H., Juillard, M., Mihoubi, F., Perendia, G., Ratto, M. and Villemot S.: Dynare: Reference Manual, Version 4, *Dynare Working Papers,* 1, Cepremap, 2011.

[2] Guerrieri, L. and Iacoviello, M.: Occbin: A Toolkit to Solve Models with Occasionally Binding Borrowing Constraints Easily, *Working paper*, January, 2014. Available at: `https://www2.bc.edu/matteo-iacoviello/research_files/TOOLKIT_PAPER.pdf`

[3] Hloušek M.: Zero lower bound on interest rate: application of DSGE model on the Czech economy: Appendix, May 2014, Available at: `www.econ.muni.cz/~hlousek/papers/zlb_appendix.pdf`.

[4] Justiniano, A. and Preston, B.: Small Open Economy DSGE models: Specification, Estimation and Model Fit. *Working Paper*, Department of Economics, Columbia University, 2004.

[5] Slanicay, M.: Structural Differences and Asymmetric Shocks between the Czech Economy and the Euro Area 12. *Review of Economic Perspectives*, Brno: Masaryk University, Vol. 11, No. 3, 2011, pp. 168-191.

Markov chain model used for sensitivity analysis of paid/unpaid claims in after-payment-due process

Jiří Hofman[1], Ladislav Lukáš[2]

Abstract. The paper uses existing Markov chain theory to analyze problem of claim payment delayed process, which serves to estimate a distribution of paid/unpaid claims after payment due. Payment condition patterns and timing of claim payments play significant role in financial management. First, the corresponding data reported in usual accounting reports are extracted to yield records of delayed payment structures. Data processing procedure issues estimations of transitions probabilities of absorption Markov chain with several transition states and two absorption ones representing paid and unpaid claims, respectively. The transition probability matrix of such absorption MC, which has a special pattern in its canonical form, enables symbolic processing of the fundamental matrix. General form of this matrix is presented in detail. Further, it is used for several numerical case studies, which are focused on sensitivity analysis of distribution of paid/unpaid claims on after payment due processes. All computations both symbolic and numerical ones are performed by sw Mathematica.

Keywords: accounts receivable analysis, fundamental matrix, absorption Markov chains, sensitivity analysis.

JEL Classification: C65, G35
AMS Classification: 60J20, 91G50

1 Introduction

Financial management in any company and cash flow management in particular, are vital to its health. Accounts receivable are the amounts owed to a business by its customers, and are comprised of a potentially large number of invoiced amounts. Accounts receivable constitute the primary source of incoming cash flow for most businesses. Gross amount of receivables and the allowance for doubtful accounts should be reported by accounting reports in detail.

Payment condition patterns and timing of claim payments play significant role in financial management. First, the corresponding data reported in usual accounting reports are extracted to yield records of delayed payment structures. Departments will also need to provide the necessary uncollectible account information in order to prepare the necessary accounting entries for prospective reserves and write-offs.

A schedule is prepared in which customer balances are classified by the length of time they have been unpaid. Because of its emphasis on time, this schedule is often called an aging schedule. As usual, the analysis of it is called aging the accounts receivable within well-established accounting framework, whereas in stochastic framework it might be called after-payment-due process analysis, as well. In general, financial management of company ought to establish a acceptable percentage relationship between the amount of receivables and expected losses from uncollectible accounts. Hence, an estimation of such relation between collected and lost payments is very significant indeed, and it plays very important role in company life cycle. Especially, when the real economy slips into recession, business faces an additional risk of customers running into financial difficulty and becoming unable to pay invoices, which all together can push a company over the edge.

After the accounts are arranged by age, the expected bad debt losses are to be determined. The most promising instrument for such task is application of absorption Markov chains. The sensitivity analysis of average amount of delayed claims paid stands in focus of financial management, too, since it enables to estimate effects of various after due diligence collection efforts.

However, usage of Markov chains for accounts receivable analysis is not new. Standard textbook of operations research and quantitative methods in management cover the topic usually, like Anderson et al [1] and Render et al [4]. The main topic in framework of existing absorption Markov chain theory concerns with detailed

[1] Univ. of West Bohemia, Faculty of Econ., Dept. of Business Admin. and Mngmt., Husova 11, 30614 Pilsen, hofman@kpm.zcu.cz.
[2] Univ. of West Bohemia, Faculty of Econ., Dpt. of Econ. and Quantitative Methods, Husova 11, 30614 Pilsen, lukasl@kem.zcu.cz.

analysis of claim payment delayed process. This one serves to estimate distribution of paid/unpaid claims after payment due, and also to calculate average amount of delayed claims paid. Sensitivity analysis presented in the paper adopts an absorption MC example excerpted from Lukáš [3], while for MC theory we refer to Yin et al [6]. Other aspects of delayed payment of claims are further discussed in Garmichael et al [2], in particular from contractor's point of view. In Sopranzetti [5], links between selling accounts receivable and underinvestment problem is thoroughly discussed.

The paper is organized as follows. After a brief introduction, the second section gives analysis of delayed payments. General item structure of suitable accounting report is sketched, and some results presenting distributions of numbers and volumes of delayed payments are given, too. The third section is split into two parts. The first one brings a general analytic form of fundamental matrix of absorption MC suited for analysis of after-payment-due process. The second one presents results of our sensitivity analysis concerning changes of average amount of delayed claims paid on transition probability variations. Finally, the fourth section brings briefly our conclusions

2 Analysis of delayed payments

First, the corresponding data reported in usual accounting reports are extracted to yield records of delayed payment structures. Our Java application reads the MS-Excel report and applies filter defined to yield a problem-oriented database of delayed payments. Usual item structure of such accounting reports accepted is following

{<recordNum>, <customer>, <invoiceNum>, <invoiceOrigCurrency>, <invoiceLocalCurrency>, <paidOrigCurr>, <paidLocCurr>, <balanceOrigCurr>, <balanceLocCurr>, <balanceTot>, <invoiceDate>, <dueDate>, <paidDate>}.

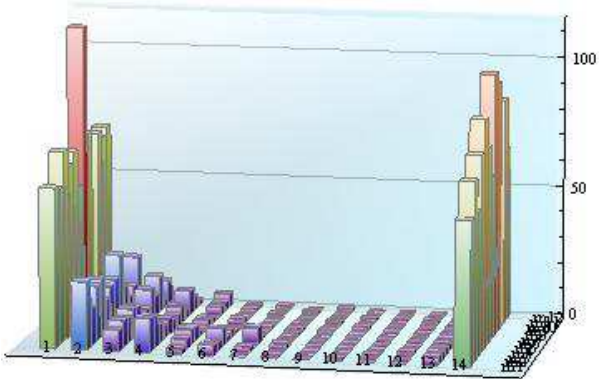
In general, a usual filter skips all records with specified payment pattern, as usual such ones with no payment delay, and extracts other records, i.e. those ones having <paidDate> greater than <dueDate>. Hence, we get a filtered sequence of records denoted D_0 which has a following structure $D_0 = \{m_k, {}_d d_k, {}_p d_k\}$, k=1,..,K where m_k gives payment in CZK, i.e. <balanceTot>, called also receivable account in accounting science. Further, ${}_d d_k$ and ${}_p d_k$ give <dueDate> and <paidDate>, respectively, and K is total number of filtered records. Finally, the delays are calculated as a 'time distance' $\delta_k = |{}_p d_k - {}_d d_k|_t$ just using the calendar. So, the D_0 is simply transformed into a new set denoted $D_1 = \{m_k, \delta_k\}$, k=1,..,K which is already well-adapted for subsequent analysis of delayed payments by Markov chain approach.

For illustration purpose, we have selected a typical accounting report provided us by particular SME in West Bohemian region for its accounts receivable analysis of whole year 2011. Using aforementioned data processing procedure, we have created both data set D_0 and D_1. In addition, we also register accumulated number of due date paid invoices and their corresponding total balance during each month.

We adopt usual bucket length of $\Delta = 30$ days for sorting our data sets. Hence, we introduce states

$s_i = [(i-1)\Delta+1, i\Delta]$, i = 1,..,12, $s_{13} = [i\Delta+1, +\infty[$, and s_{14} for due date paid invoices.

First on Fig. 1 and Fig. 2, we show number of pending payments and their balances per month during whole year 2011, respectively, together with corresponding due date paid invoices per the month, too.



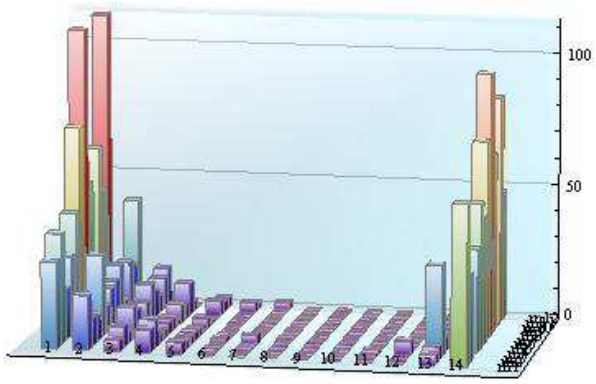


Figure 1 number of pending payments – own results **Figure 2** balance of pending payments – own results

Horizontal axes represent the states s_i, i = 1,..,14, whereas width axes represent months of year 2011. The volumes of unpaid balances per month depicted on Fig. 2 are scaled by 10^5 just in order to keep them within proper bounds. Since the company financial management concerns mostly on timeliness and volumes of payments, the distribution plots on Fig. 1 and Fig. 2 are created by back-lagging approach. It means that quantities of primary

interest are ${}_pd_k$ which are gathered into month long buckets, too. These ones point uniquely to their corresponding ${}_dd_k$ and m_k, or to their delays δ_k respectively, which are subsequently sorted into states s_i. Inspecting Fig. 1 and Fig. 2 we simply conclude that states s_1,.., s_4 play very important role within delay payment structure.

In particular, the figures Fig. 3 – Fig. 6 show results which correspond to states s_1, s_2, s_3, and s_4. They are selected particularly, as the overall majority of delayed claim numbers and balance amounts are concentrated therein. Since all these four figures are plotted in a unique way thus keeping colors/shades applied to be the same, we may describe in detail just one figure of them – let make it with Fig. 3, and the others, i.e. Fig. 4 – Fig. 6 just follow the same legends.

The Fig. 3 displays two curves over the year 2011. The first curve being identifiable as a line with one sharp peak only represents number of states s_1 caught by each month of year particularly. The second curve being identifiable as a line with two peaks represents volumes m_k scaled by 10^5 of pending balances corresponding to the same states s_1. Thus both curve identifications being given by the Fig.3, we can simply decode each couple of curves and their corresponding meanings on the remaining figures, i.e. Fig. 4, Fig. 5, and Fig. 6, as well.

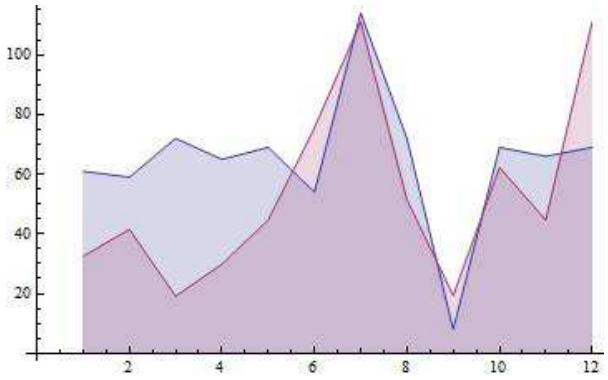


Figure 3 number of s_1 and their m_k – own results

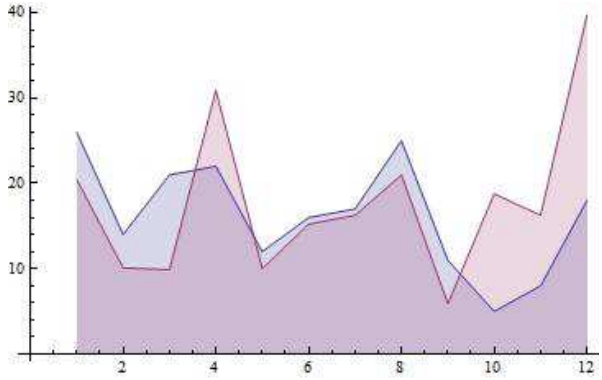


Figure 4 number of s_2 and their m_k – own results

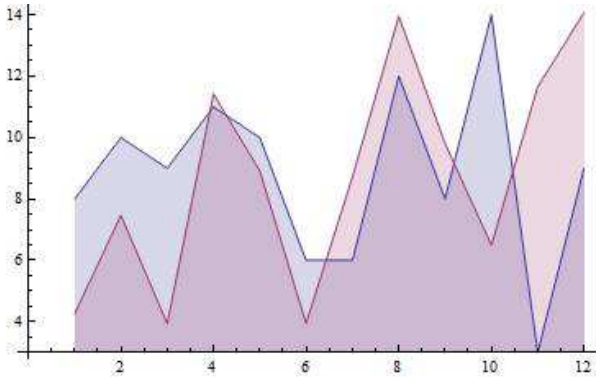


Figure 5 number of s_3 and their m_k – own results

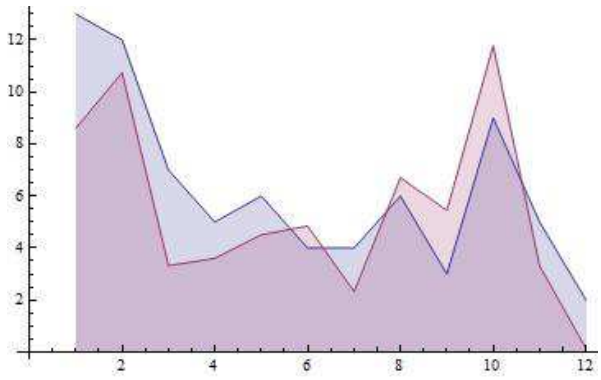


Figure 6 number of s_4 and their m_k – own results

Such results provide sufficient material for construction of transition probability matrix **P** for proper absorption MC. Main purpose of the MC analysis knowing matrix **P** is to estimate an amount of paid/unpaid receivables with delayed due date payments.

3 Absorption Markov chains and sensitivity analysis

For illustration purposes and for sensitivity analysis, we select a simpler example given in Lukáš [3], pp. 48 – 51. Total number of states introduced is five, as three of them, s_1, s_2, and s_3, stand for transient states thus representing usual 30-days long buckets for collecting delayed claims. States s_4 and s_5 stand for absorption states presenting paid and unpaid/lost receivables. The corresponding problem-oriented database being provided, we may create the transition probability matrix describing behavior of delayed claims. Next, this matrix is permuted to get well-established canonical form $\mathbf{P} = [p_{ij}]$, $i,j=1,..,5$ as follows

$$\begin{matrix} s_1 & s_2 & s_3 & s_4 & s_5 \end{matrix}$$
$$\begin{bmatrix} 0 & 0.77 & 0 & 0.23 & 0 \\ 0 & 0 & 0.34 & 0.66 & 0 \\ 0 & 0 & 0 & 0.73 & 0.27 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \begin{matrix} s_4 & s_5 & s_1 & s_2 & s_3 \end{matrix}$$
$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0.23 & 0 & 0 & 0.77 & 0 \\ 0.66 & 0 & 0 & 0 & 0.34 \\ 0.73 & 0.27 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}. \qquad (1)$$

As well-known, the crucial notion for any absorption MC is its fundamental matrix $\mathbf{N} = (\mathbf{I}–\mathbf{Q})^{-1} = \sum_{n=0}^{\infty} \mathbf{Q}^{\boldsymbol{n}}$. It enables to calculate average amount of paid and unpaid claims, called also uncollectible receivable accounts, from outstanding ones. Let denote **t** a vector of reported balances belonging to transient states, then $\mathbf{y}^{\mathrm{T}} = \mathbf{t}^{\mathrm{T}}\mathbf{B}$, **B** = **N R** gives the average amount of paid and unpaid claims, which is most desirable information from financial management point of view. In our case assuming the vector **t** is given, we directly get the result **y** as follows

$$\mathbf{t}^{\mathrm{T}} = (4.030, 9.097, 3.377)\ [\mathrm{MCZK}], \qquad \mathbf{y}^{\mathrm{T}} = (14.472240, 2.031760)\ [\mathrm{MCZK}]. \qquad (2)$$

3.1 General form of fundamental matrix

There is not a problem to show that assuming matrices **Q** and **R** used in (1) in general form, i.e. **Q** containing non-zero transition probabilities $p_{3,4}$ and $p_{4,5}$ only, and **R** having $p_{3,1}$, $p_{4,1}$, $p_{5,1}$ and $p_{5,2}$ as its non-zero entries only, too, the corresponding fundamental matrix **N** is upper triangular one with the unit main diagonal. However, aiming to get a general analytic form of the matrix **N**, we simplify the notation of non-zero elements of matrix **Q** first, inducing non-zero elements of **R**, too, which yields the **N** taking the form as follows

$$\mathbf{Q} = \begin{bmatrix} 0 & p_1 & 0 \\ 0 & 0 & p_2 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1-p_1 & 0 \\ 1-p_2 & 0 \\ 1-q & \mathrm{q} \end{bmatrix}, \qquad \mathbf{N} = \begin{bmatrix} 1 & p_1 & p_1p_2 \\ 0 & 1 & p_2 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (3)$$

where p_1 and p_2 stand for $p_{3,4}$ and $p_{4,5}$ specified earlier, which express conditional probabilities of system transitions from one transient state to the next one $s_i \rightarrow s_{i+1}$, i = 1,2, respectively. Hence, the corresponding non-zero transition probabilities from transient states to absorption ones are following $P(s_4|s_i) = p_{2+i,1} = 1 – p_i$, i = 1,2, and from s_3 $P(s_4|s_3) = p_{5,1} = 1 – q$, and $P(s_5|s_3) = p_{5,2} = q$, respectively. All other entries are zero.

Practical reason, for expressing fundamental matrix **N** in general analytic form for arbitrary structure of state space assigned to absorption MC used for analysis of paid/unpaid claims, is possibility to eliminate time-consuming matrix inversion, in case the number of transient states is large.

It correspond to situation when aging structure of outstanding receivable accounts is relatively fine, for example defining the bucket length instead of 30-days just 1-day only, thus increasing the number of transient states introduced within a year from 12 ones to 360 ones. Of course, it does depend upon the size of data set D_0, or D_1, respectively, and on the quantity K, in particular.

For illustration purposes, we just show a general structure of fundamental matrix **N** for n = 5, assuming the matrix **Q** is 5x5 with its pattern induced by (3) generalized to that size

$$\mathbf{N} = \begin{bmatrix} 1 & p_1 & p_1p_2 & p_1p_2p_3 & p_1p_2p_3p_4 \\ 0 & 1 & p_2 & p_2p_3 & p_2p_3p_4 \\ 0 & 0 & 1 & p_3 & p_3p_4 \\ 0 & 0 & 0 & 1 & p_4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \qquad (4)$$

We have used symbolic calculation power of Mathematica to get the general analytic form of **N** desired for size ranging n = 3, 4, .., 9. For example taking n = 5, the corresponding snippet of Mathematica code is

```
ImQ5x5symb={{1,-p1,0,0,0},{0,1,-p2,0,0},{0,0,1,-p3,0},{0,0,0,1,-p4},
           {0,0,0,0,1}};
N5x5symb=Inverse[ImQ5x5symb]; %//FullSimplify; %//MatrixForm
R5x5symb={{1-p1,0},{1-p2,0},{1-p3,0},{1-p4,0},{1-q,q}};
N5x5symb.R5x5symb; %//FullSimplify; %//MatrixForm
```

Checking whether **N**(**I**–**Q**) = **I** = (**I**–**Q**)**N** holds is done with the following code

```
ImQ5x5symb.N5x5symb; %//FullSimplify; %//MatrixForm
N5x5symb.ImQ5x5symb; %//FullSimplify; %//MatrixForm
```

Here, array `ImQ5x5symb` contains matrix **I**–**Q** stored row-vise, and `R5x5symb` gives matrix R stored in the similar fashion. Meaning of all non-zero quantities therein is self-evident.

Results of our symbolic calculations lead us to formulate following lemma.

Lemma 1. *Let bi-diagonal* (n,n) *matrix* $\mathbf{V} = \mathbf{I}-\mathbf{Q}$ *provides decomposition* $\mathbf{V} = \mathbf{L}+\mathbf{D}+\mathbf{U}$ *where* $\mathbf{L}$ *is lower triangular null matrix,* $\mathbf{D}$ *is diagonal unit matrix, and* $\mathbf{U}$ *is upper triangular matrix taking row-wise form*

$((-p_1, 0, 0,\ldots,0), (-p_2, 0,\ldots,0), \ldots, (-p_{n-2}, 0), (-p_{n-1}))$ *with non-zero entries* p_i, $i=1,..,n-1$

then matrix $\mathbf{N} = \mathbf{V}^{-1}$ *has unique decomposition* $\mathbf{N} = \mathbf{L}+\mathbf{D}+\mathbf{W}$ *where* $\mathbf{W} = [w_{ij}]$ *is full upper triangular matrix with elements* $w_{ij} = \prod_{k=i}^{j-1} p_k$, $i=1,..,n-1$, $j=i+1,..,n$.

Quasi-Proof. The symbolic calculation power of Mathematica was also used to verify the lemma for all matrix orders up to $n = 9$. This ends our quasi-proof.

Note. Since we are concentrated on computational solution of accounts receivable problems mainly, we do not provide here a rigorous proof of the lemma using advanced matrix algebra together with mathematical induction tool in general case, as it has been mentioned by anonymous referee, however, without any reference given us thereon.

3.2 Sensitivity analysis

Finally, we describe some numerical results regarding sensitivity analysis of vector **y** upon slight variations of probabilities p_1, p_2, and q, respectively. The basic data are the same as provided by (1) and (2).

General sensitivity analysis in full scope of such prospective 3-D variations is much more involved, and even more it prevents any possibility to plot results. Hence, we restrict ourselves by one-parametric variations only, which enable us to plot the results clearly. First, we modify the matrices **Q** and **R** as to be one variation parameter ρ dependent only, denoted $\mathbf{Q}(\rho)$ and $\mathbf{R}(\rho)$. The variations of probabilities take following form

$$p_1(\rho) = p_1 - \rho, \quad p_2(\rho) = p_2 - \rho, \quad q(\rho) = q - \rho, \quad \rho = [0.01, 0.1] \text{ by step } 0.01, \tag{5}$$

which interpretation is rather simple, since in each particular case it represents an increase of complement probability of absorption state s_4 by one percent, knowing that s_4 represents paid claims after their due date, in general. Hence, from company financial management point of view, such analysis provides insight to diligent activity effects for collecting outstanding receivables, and it is most requested and desirable one, too.

In detail, the matrices $\mathbf{Q}(\rho)$ and $\mathbf{R}(\rho)$ are formed both using (4) for variation of one selected probability only, and following (1) ceteris paribus. Snippet of Mathematica code realizing such computations is following

```
Nmtx={{1,p1,p1 p2},{0,1,p2},{0,0,1}}; Rmtx={{1-p1,0},{1-p2,0},{1-q,q}};
Bmtx=Nmtx.Rmtx;
tvect={t1,t2,t3}; yvect=tvect.Bmtx;
yvectB=yvect/.{p1→.77,p2→.34,q→.27,t1→4.030,t2→9.097,t3→3.377};
scl=.01;
ex4sensitp1p2q=Table[△paid=scl i;
   wp1=(yvect/.{p1→(.77-△paid),p2→.34,q→.27,t1→4.030,t2→9.097,t3→3.377})-
yvectB;
   wp2=(yvect/.{p1→.77,p2→(.34-△paid),q→.27,t1→4.030,t2→9.097,t3→3.377})-
yvectB;
   wq=(yvect/.{p1→.77,p2→.34,q→(.27-△paid),t1→4.030,t2→9.097,t3→3.377})
-yvectB;
   {wp1[[1]],wp2[[1]],wq[[1]]},{i,1,10}];
```

where array `ex4sensitp1p2q` contains all results plotted on Fig. 7. The arrays `Nmtx, Rmtx` and `Bmtx` contain matrices **N**, **R**, and **B**, respectively, all in symbolic form with variables `p1`, `p2`, and `q` defined. The most important quantity we desire is vector **y** which is represented by array `yvect`. The vector **y** computed for already given values of outstanding claims as being reported by balances belonging to transient states by vector **t** is calculated by Mathematica simply using its specific assignment rule, and it is represented by array `yvectB`.

Now, let discuss briefly the results depicted on Fig. 7. The first row of vertical bars thereon displays changes of the first element of vector **y** which gives increases of average paid claims in [MCZK] belonging to state s_4, assuming $p_1(\rho)$ being varied by (4) in ten grades in total, stepping one percent up each time.

The second row gives analogous gains, however in case of varying $p_2(\rho)$, which looks definitely better, thus concluding that prospective diligence effort for collecting delayed claims is to bring more effects, hence it is worth to be concentrated thereupon.

Finally, the rear row depicts gains again. Sure, since the overall balance between vectors **y** and **t** must always hold, the gains in average amount of delay claims paid are the same as equivalent decreases of average unpaid claims in [MCZK] as well, which is instructive, as well. It also shows that the most effective diligence effort should be concentrated upon decreasing transition probability q measuring transition from the last transient state s_3 into the unpaid state s_5, i.e. $P(s_5|s_3) = q$ in particular, as much as possible.

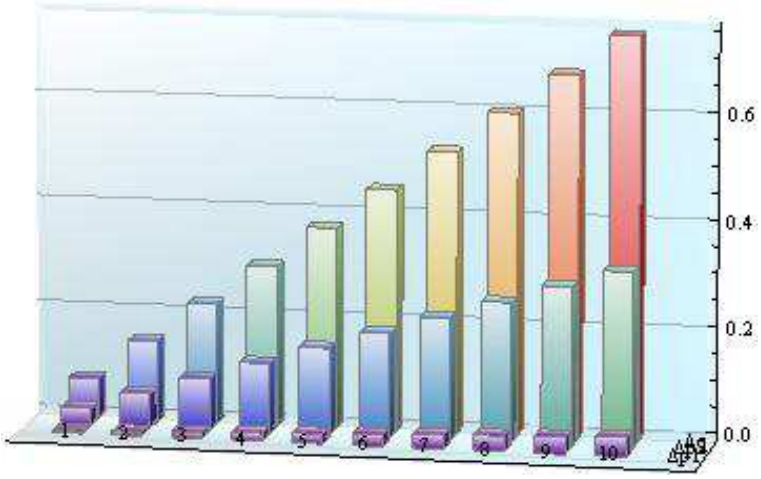


Figure 7 sensitivity analysis of **y**[1] using **Q**(ρ) and **R**(ρ) – own results

4 Conclusion

In our paper, we have discussed briefly links between item structure of accounting reports and problem-oriented database properly suited for analysis of due date delayed claims, first. Second, we have presented general analytic form of fundamental matrix of any absorption Markov chain used for after-payment-due process analysis. Finally, we have shown some interesting results regarding sensitivity analysis of dominantly considered average amount of paid delayed claims on one-parametric variation of transition probabilities from all transient states given. Core snippets of Mathematica code are presented as well. Ongoing research will be focused on two challenging topics:

- further development of sensitivity analysis within the framework of collection possibilities of outstanding claims and its implementation in Mathematica;
- connection of account receivable advanced analysis with credit risk procedures.

Acknowledgements

The research project was supported by Grant SGS-2014-047 of the Faculty of Economics, University of West Bohemia in Pilsen.

References

[1] Anderson, D.R., Sweeney, D.J., and Williams, T.A.: *An Introduction to Management Science – Quantitative Approches to Decision Making*. 5-th Ed., West Publ. Co., St. Paul, Minnesota, 1988, 688-693.

[2] Garmichael, D.G., and Balatbat, M.C.A.: A contractor's analysis of the likelihood of payment of claims. *Journal of Financial Management of Property and Construction* **15**, 2 (2010), 102–117.

[3] Lukáš, L.: *Probabilistic models in management – Markov chains and waiting line systems*. (in czech: Pravděpodobnostní modely v managementu – Markovovy řetězce a systémy hromadné obsluhy) Academia, Prague, 2009.

[4] Render, B., Stair, M.R.Jr., and Hanna, M.E.: *Quantitative Analysis for Management*. 8-th Ed., Pearson Education, Prentica Hall, Upper Saddle River, New Jersey, 2003, 655-659.

[5] Sopranzetti, B.J.: Selling accounts receivable and the underinvestment problem. *The Quarterly Review of Econ. and Fin.* **39**, 2 (1999), 291–301.

[6] Yin, G., Zhang, Q.: *Discrete-time Markov Chains*. Springer, New York, NY, 2005.

Multiple-Criteria Fuzzy Evaluation in FuzzME - Transitions Between Different Aggregation Operators

Pavel Holeček[1], Jana Talašová[2]

Abstract. The FuzzME is a software tool for multiple-criteria fuzzy evaluation. The type of evaluation employed in the fuzzy models fully agrees with the paradigm of the fuzzy set theory; the evaluations express the (fuzzy) degrees of fulfillment of the corresponding goals. The basic structure used for the evaluation is called a goals tree. Within the goals tree, the aggregation of partial fuzzy evaluations is done either by one of fuzzified aggregation operators or by a fuzzy expert system. In the FuzzME, the following aggregation methods are supported: fuzzy weighted average, fuzzy OWA operator, fuzzified WOWA operator, fuzzified discrete Choquet integral, and fuzzy expert system. The comprehensive description of the methods and their use in FuzzME has been published in [3].

The paper contains a brief description of the methods. The main part of the paper studies the situation when it turns out that the relationship among the criteria is more complex than it was expected. In this case, it would be favorable to replace the original aggregation operator with a more general one with a minimum effort. For this purpose, two algorithms will be presented. The first one allows to derive FNV-measure (fuzzy-number-valued fuzzy measure) for the fuzzy Choquet integral from the parameters of the original aggregation operator (any of the above mentioned aggregation operators can be used). The second algorithm can be used to derive a fuzzy rule base so that the result would be as similar as possible to the original aggregation method.

Keywords: Multiple-criteria evaluation, aggregation operators, software

JEL classification: C44
AMS classification: 91B06

1 Introduction

In practice, multiple-criteria decision-making problems are quite common – e.g. selecting the best candidate when a company needs to hire a new employee. In the real-world applications, we have to often work with an expert evaluation (e.g. evaluation of the stress tolerance, or ability to work in a team), where some uncertainty is always present. Also, the values of quantitative criteria need not to be always given preciously. Because of the complexity of some problems (number of criteria and dependencies among them), it is advantageous to decompose the multiple-criteria evaluation problem into a tree structure.

One of the fuzzy methods of multiple-criteria evaluation that can work with uncertain values of criteria and employs a tree structure is a generalized partial goals method [3, 5], which will be used in this paper. The paper is organized as follows. Some basic notions of the fuzzy sets theory are mentioned first. Next, the generalized partial goals method and its software implementation (the FuzzME software) are described briefly. One of the advantages of the method is that various aggregation operators can be used and combined in a single model. The main part of the paper then deals with the situation when the model has already been designed but it has turned out that another aggregation operator would be more suitable. As it will be shown later, these situations can occur quite easily. The paper gives a recommendation, what could be done in those cases in order to make the change as easy as possible for the

[1]Palacký University in Olomouc, Faculty of Science, Department. of Mathematical Analysis and Applications of Mathematics, pavel.holecek@upol.cz
[2]Palacký University in Olomouc, Faculty of Science, Department. of Mathematical Analysis and Applications of Mathematics, jana.talasova@upol.cz

decision-maker. Two algorithms will be proposed and the advantages of their use will be demonstrated on examples.

2 Basics of the fuzzy sets theory

A fuzzy set A on a universal set X is characterized by its membership function $A : X \to [0,1]$. $Ker\ A$ denotes a kernel of A, $Ker\ A = \{x \in X \mid A(x) = 1\}$. For any $\alpha \in [0,1]$, A_α denotes an α-cut of A, $A_\alpha = \{x \in X \mid A(x) \geq \alpha\}$. A support of A is defined as $Supp\ A = \{x \in X \mid A(x) > 0\}$.

A fuzzy number is a fuzzy set C on the set of all real numbers $\Re$ which satisfies the following conditions: a) the kernel of C, $Ker\ C$, is not empty, b) the α-cuts of C, C_α, are closed intervals for all $\alpha \in (0,1]$, c) the support of C, $Supp\ C$, is bounded. A fuzzy number C is called to be defined on $[a,b]$, if $Supp\ C \subseteq [a,b]$. The set of all fuzzy numbers on $[a,b]$ will be denoted by the symbol $\mathcal{F}_N([a,b])$. Real numbers $c^1 \leq c^2 \leq c^3 \leq c^4$ are called significant values of the fuzzy number C if the following holds: $[c^1,c^4] = Cl(Supp\ C), [c^2,c^3] = Ker\ C$, where $Cl(Supp\ C)$ denotes a closure of $Supp\ C$.

Any fuzzy number C can be characterized by a pair of functions $\underline{c} : [0,1] \to \Re$, $\overline{c} : [0,1] \to \Re$ which are defined by the following formulas: $C_\alpha = [\underline{c}(\alpha), \overline{c}(\alpha)]$ for all $\alpha \in (0,1]$, and $Cl(Supp\ C) = [\underline{c}(0), \overline{c}(0)]$. The fuzzy number C is called to be linear if both the functions $\underline{c}$, $\overline{c}$ are linear.

A linear fuzzy number C is called triangular if $c^2 = c^3$, otherwise it is called trapezoidal. In the examples, linear fuzzy numbers will be described by their significant points. A linear fuzzy numbers C will be defined using the following notation: $C = (c^1, c^2, c^3, c^4)$ if C is trapezoidal, or simply $C = (c^1, c^2, c^4)$ if C is a triangular fuzzy number.

The fuzzy set $\tilde{0}$ will represent the crisp 0 in form of a fuzzy singleton, i.e. $\tilde{0}(0) = 1$ and $\tilde{0}(x) = 0$, for $x \in \Re$, $x \neq 0$. Similarly, the fuzzy set $\tilde{1}$ will be a fuzzy singleton representing the crisp 1, i.e. $\tilde{1}(1) = 1$ and $\tilde{1}(x) = 0$, for $x \in \Re$, $x \neq 1$.

A linguistic variable is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values, X is a universal set on which the meanings of the linguistic values are defined, G is a syntactic rule for generating values in $\mathcal{T}(\mathcal{V})$, and M is a semantic rule which maps each linguistic value $\mathcal{C} \in \mathcal{T}(\mathcal{V})$ to its mathematical meaning, $C = M(\mathcal{C})$, which is a fuzzy set on X. In this paper, the linguistic values (also called *terms*) and their mathematical meanings (fuzzy numbers) are distinguished by means of a different font.

An ordering of fuzzy numbers is defined as follows: a fuzzy number C is greater than or equal to a fuzzy number D, if $C_\alpha \geq D_\alpha$ for all $\alpha \in (0,1]$. The inequality of the α-cuts $C_\alpha \geq D_\alpha$ is the inequality of intervals $C_\alpha = [\underline{c}(\alpha), \overline{c}(\alpha)]$, $D_\alpha = [\underline{d}(\alpha), \overline{d}(\alpha)]$ which is defined as

$$[\underline{c}(\alpha), \overline{c}(\alpha)] \geq [\underline{d}(\alpha), \overline{d}(\alpha)] \text{ if, and only if, } \underline{c}(\alpha) \geq \underline{d}(\alpha) \text{ and } \overline{c}(\alpha) \geq \overline{d}(\alpha).$$

Let $G = \{G_1, \ldots, G_m\}$ be a nonempty finite set, $\wp(G)$ be the family of all its subsets. Then, a FNV-fuzzy measure on G is a set function $\tilde{\mu} : \wp(G) \to \mathcal{F}_N([0,1])$ satisfying the following conditions:

- $\tilde{\mu}(\emptyset) = \tilde{0}$, $\tilde{\mu}(G) = \tilde{1}$, and
- $C \subseteq D$ implies $\tilde{\mu}(C) \leq \tilde{\mu}(D)$ for any $C, D \in \wp(G)$.

A fuzzy rule base is a set of if-then rules, which are used to model the relation between the values of the input variables and the output variable, and it is of the following form:

$$\begin{aligned}
&\text{If } \mathcal{E}_1 \text{ is } \mathcal{U}_{1,1} \text{ and } \ldots \text{ and } \mathcal{E}_m \text{ is } \mathcal{U}_{1,m}, \text{then } \mathcal{E} \text{ is } \mathcal{U}_1,\\
&\text{If } \mathcal{E}_1 \text{ is } \mathcal{U}_{2,1} \text{ and } \ldots \text{ and } \mathcal{E}_m \text{ is } \mathcal{U}_{2,m}, \text{then } \mathcal{E} \text{ is } \mathcal{U}_2,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\text{If } \mathcal{E}_1 \text{ is } \mathcal{U}_{n,1} \text{ and } \ldots \text{ and } \mathcal{E}_m \text{ is } \mathcal{U}_{n,m}, \text{then } \mathcal{E} \text{ is } \mathcal{U}_n,
\end{aligned}$$

where for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$: $(\mathcal{E}_j, \mathcal{T}(\mathcal{E}_j), [a_j, b_j], M_j, G_j)$ are linguistic variables for the inputs, $\mathcal{U}_{ij} \in \mathcal{T}(\mathcal{E}_j)$ are their linguistic values, $U_{ij} = M_j(\mathcal{U}_{ij})$ are fuzzy numbers representing their meanings, and similarly, $(\mathcal{E}, \mathcal{T}(\mathcal{E}), [a,b], M, G)$ is the linguistic variable representing the output, $\mathcal{U}_i \in \mathcal{T}(\mathcal{E})$ are its linguistic values, $U_i = M(\mathcal{U}_i)$ are fuzzy numbers representing their meanings.

More detailed information on the fuzzy set theory can be found for example in [1].

3 Generalized partial goals method

This paper deals with a generalization of the partial goals method [5, 3]. The type of the used evaluation fully agrees with the paradigm of the fuzzy set theory. The evaluations express the (fuzzy) degrees of fulfillment of the corresponding goals. All evaluations are therefore fuzzy numbers on $[0, 1]$, where $\tilde{0}$ means that the alternative is completely unsatisfactory according to the particular partial goal and, on the other side, $\tilde{1}$ means that we are fully satisfied with the alternative according to this partial goal. The partial goals are organized into a hierarchical tree structure that is called a goals tree.

The partial goals at the ends of the goals tree branches are connected to criteria, which can be either qualitative or quantitative. For qualitative criteria, linguistic variables have to be defined in advance. When an alternative is evaluated with respect to such a criterion, the expert selects the best fitting term from the linguistic variable. For quantitative criteria, an evaluation function (membership function of the corresponding partial goal) has to be defined. Then the evaluation is calculated from the measured value of the criterion (crisp or fuzzy number) using the extension principle [5].

Evaluation of the partial goals on a higher level is obtained from the evaluations of the partial goals on the lower level by one of the aggregation operators or by a fuzzy expert system. The following aggregation operators are supported: fuzzy weighted average, fuzzy OWA operator, fuzzified WOWA operator, and fuzzified discrete Choquet integral. The evaluation of the main goal (the root node of the goals tree) is then the overall evaluation of the alternative.

Each of the mentioned aggregation operators can be used under different conditions. Fuzzy weighted average and fuzzy OWA operator are suitable only when the aggregated criteria are independent. Fuzzy Choquet integral can be used, if there is a relationship of complementarity or redundancy among the criteria. And finally, a fuzzy expert system can be used even in cases when there are complex relationships among the aggregated criteria. The more general the aggregation operator is, the more information is required to be set by the expert. Therefore, it is always advantageous to choose the simplest mean of aggregation that can be used for the particular situation.

As it has been already mentioned, all evaluations, including the overall evaluation of the alternative, are expressed by fuzzy numbers on the interval $[0, 1]$. This presents a problem with a direct comparison of the evaluations. However, any fuzzy numbers can be compared by their centers of gravity [1].

Because of the space limitation, the reader is kindly asked to refer to the book chapter [3] for more information. The mentioned chapter contains a comprehensive description of the method and comparison to some other methods.

4 FuzzME software

Many methods of multiple criteria evaluation require non-trivial calculations. It is thus quite natural that the fuzzy versions of those methods are computationally even more demanding. To use those methods in practice, a software tool is therefore necessary.

The FuzzME software has been created at the Palacky University in Olomouc. It implements the methods mentioned in this paper in a user-friendly way. It allows to design a goals tree for the given problem and to use it for evaluation of a set of alternatives. The alternatives together with their criteria values can be imported from Excel and the results of the evaluation can be also exported to the Excel for further processing. The FuzzME strives to make the whole evaluation process as easy as possible. The results are presented not only in a numerical form but also graphically and verbally.

Of course, a Matlab code could be created and used for all of these methods with some effort. However, the great advantage of a special software tool over the Matlab code is that it makes it possible to see the impact of any changes in the model in the real time. When a weight or some other parameter of the model is changed, the FuzzME recalculates all of the evaluations immediately. This, together with other functions, allows the expert to have a good grasp of the created model behavior.

A demo version of the FuzzME software is available on *http://www.fuzzme.net*. The readers can therefore try the mentioned method on their own. Also, the transitions between the aggregation operator, which are the main topic of this paper, have already been implemented into the software.

5 Transitions between different aggregation operators

In the practice, the creation of the evaluation model is seldom a one-step process. The model needs to be tested and, as the result of these tests, its parameters can be adjusted. It can easily turn out, that another type of aggregation would be more suitable for a particular partial goal. The expert could have either underestimated the complexity of the relationship among the partial goals or the client's requirements could have changed. One way or another, the use of a more complex aggregation method could be necessary.

The main goal of the paper is to give some recommendation what should be done in those cases. For example, a fuzzy Choquet integral is used and it is realized that for some combinations of criteria values the result should be different and a fuzzy expert system would be more suitable. The FNV-fuzzy measure for the Choquet integral can be then used to derive the fuzzy rule base automatically. Then, the expert just changes a few rules for the criteria values of interest, instead of creating the whole rule base from the beginning. A lot of long and tedious work can be saved this way. Similarly, it could be possible that the expert needs to use fuzzy Choquet integral instead of a fuzzy weighted average. It would be advantageous, if the (additive) FNV-fuzzy measure could be derived from the original normalized fuzzy weights [4] and the expert would just adjust a few of its values instead of setting all of them.

In this paper, those two types of transitions will be studied in detail. In the first one, the original aggregation operator is replaced with a fuzzy Choquet integral. In the latter one, a fuzzy expert system is used as the new aggregation method for the particular partial goal.

In the further text, $m \in \mathbb{N}$ will denote the number of the aggregated evaluations (i.e. the number of children of the particular goals tree node) and the FNV-function (fuzzy-number-valued function) f, $f : \mathcal{F}_N([0,1])^m \to \mathcal{F}_N([0,1])$, will be the original function that has been used for the aggregation (i.e. a fuzzy weighted average with some normalized fuzzy weights).

5.1 Using the fuzzy Choquet integral

In this section, we will study the case when a fuzzy weighted average, a fuzzy OWA or a fuzzified WOWA has been used originally for the aggregation and the expert would like to use the fuzzy Choquet integral instead. For the fuzzy Choquet integral, a FNV-fuzzy measure has to be set. This normally means to set $2^m - 2$ values of the measure (which are fuzzy numbers). The following algorithm can be used to create the FNV-fuzzy measure automatically so that the result of the aggregation would be the same as in case of the original aggregation operator. This FNV-fuzzy measure can be then adjusted by the expert. The algorithm is based on a method of setting a fuzzy measure by an expert in the crisp case, which is presented in [2]

Algorithm 1. The FNV-fuzzy measure $\tilde{\mu}$ is constructed as follows:

1. Set $\tilde{\mu}(\emptyset) = \tilde{0}$ by the definition,
2. Set $\tilde{\mu}(G) = \tilde{1}$ by the definition.
3. For the rest of the $2^m - 2$ values, the measure of A, $A \subset G$, is calculated as $\tilde{\mu}(A) = f(A_1, \ldots, A_m)$, where

$$A_i = \begin{cases} \tilde{1} & \text{if } G_i \in A \\ \tilde{0} & \text{otherwise} \end{cases}$$

Example

Let us assume that high school students need to be evaluated according to their performance in Mathematics, Physics and English. The expert used a fuzzy weighted average originally. The subjects had the following normalized fuzzy weights [4]: $W_{Mathematics} = (0.3, 0.4, 0.5)$, $W_{Physics} = (0.25, 0.35, 0.45)$, and $W_{English} = (0.15, 0.25, 0.35)$. However, in this case, the fuzzy weighted average would be inappropriate. The reason is that there is a relationship of redundancy between the Mathematics and the Physics. These two partial goals are overlapping. There are some common pieces of knowledge that are necessary in order to succeed in either of them. Therefore a fuzzy Choquet integral should be used. The Algorithm 1 will be used to generate the new FNV-fuzzy measure $\tilde{\mu}$, whose values are:

- $\tilde{\mu}(\emptyset) = \tilde{0}$
- $\tilde{\mu}(Mathematics) = (0.3, 0.4, 0.5)$
- $\tilde{\mu}(Physics) = (0.25, 0.35, 0.45)$
- $\tilde{\mu}(English) = (0.15, 0.25, 0.35)$
- $\tilde{\mu}(Mathematics, Physics) = (0.65, 0.75, 0.85)$
- $\tilde{\mu}(Mathematics, English) = (0.55, 0.65, 0.75)$
- $\tilde{\mu}(Physics, English) = (0.5, 0.6, 0.7)$
- $\tilde{\mu}(Mathematics, Physics, English) = \tilde{1}$

Because the relationship of redundancy is present only between the Mathematics and the Physics, the expert needs to modify just one single value to form the final FNV-fuzzy measure. The expert could decrease it for example to $(0.5, 0.6, 0.7)$.

In this particular example, the expert had to adjust just one measure value instead of setting 6 of them (the other 2 measure values are given by the FNV-fuzzy measure definition).

5.2 Using a fuzzy expert system

The Choquet integral can handle only some types of criteria interactions. If the relationship among the criteria is more complex, a fuzzy expert system can be used for the evaluation. In this section, we will assume that a fuzzy weighted average, a fuzzy OWA operator, a fuzzified WOWA operator or a fuzzy Choquet integral has been originally used for the evaluation and the expert wants to use a fuzzy expert system instead. The problem that will be tackled is to create a fuzzy rule base so that the result would be as similar to the results of the original aggregation operator as possible. For this purpose, the following algorithm can be used.

Algorithm 2. For any possible combination of the criteria values (terms of the corresponding linguistic variables), a rule is created. Let s_i, $i = 1, \ldots, m$, denote the number of the terms of the linguistic variable $\mathcal{E}_i$ and let s denote the number of the terms of the linguistic variable $\mathcal{E}$. Then $n = s_1 \cdot s_2 \cdot \ \cdots \ \cdot s_m$ denotes the total number of the rules that should be created. The following steps are performed for each of them. Let the antecedent of such an i-th rule, $i = 1, \ldots, n$, be

$$\text{If } \mathcal{E}_1 \text{ is } \mathcal{U}_{i,1} \text{ and } \ldots \text{ and } \mathcal{E}_m \text{ is } \mathcal{U}_{i,m}.$$

The consequent $\mathcal{U}_i$ for this rule is determined in the following way:

- A fuzzy number B_i is calculated as $B_i = f(U_{i1}, \ldots, U_{im})$, where $U_{ij} = M_j(\mathcal{U}_{ij})$, $j = 1, \ldots, m$,
- The linguistic term $\mathcal{U}_i \in \mathcal{T}(\mathcal{E})$ is then found by the linguistic approximation of B_i to the linguistic variable $\mathcal{E}$, i.e. $\mathcal{U}_i$ is such a term for whose mathematical meaning, the fuzzy number U_i, $U_i = M(\mathcal{U}_i)$, it holds that $\forall \mathcal{A} \in \mathcal{T}(\mathcal{E}) : P(B_i, U_i) \geq P(B_i, A)$, where $A = M(\mathcal{A})$ and

$$P(B_i, A) = 1 - \frac{\int_a^b |\, B_i(x) - A(x) \,|\ dx}{\int_a^b (B_i(x) + A(x))\ dx}.$$

The quality of the approximation of the original aggregation method by a fuzzy rule base depends on the number of terms in the linguistic variables for the aggregated partial goals and for the overall evaluation. Because of the size of the resulting rule base, the method is suitable only when no more than 3 partial evaluations should be aggregated. For more inputs, it is recommended to split the partial goal into more partial goals with fewer inputs.

Example

Let us assume that high-school students applying for the study on a university are evaluated with respect to the entrance exam results and their grades. The fuzzy weighted average has been used and the following normalized fuzzy weights have been set: $W_{Exam} = (0.6, 0.7, 0.8)$ and $W_{Grades} = (0.2, 0.3, 0.4)$. The fuzzy weighed average performed well. However, the university placed a new requirement on the evaluation: the students who failed the entrance exam completely should be evaluated as *very bad* no matter what their grades are. This requirement cannot be fulfilled by the fuzzy weighted average any more. The new evaluation function has to be described by a fuzzy rule base.

The same linguistic variable has been used for the evaluation of the entrance exam and the grades and for the overall evaluation of the student (see Figure 1). The linguistic variable contain 5 terms, so the resulting fuzzy rule base will consist of 25 rules. Applying the Algorithm 2, the following rules are generated. Because of the space limitation, only an excerpt will be given:

1. If Entrance exam result = Very bad and Grades = Very bad, then Student evaluation = Very bad.
2. If Entrance exam result = Very bad and Grades = Bad, then Student evaluation = Very bad.
3. If Entrance exam result = Very bad and Grades = Average, then Student evaluation = Bad.
4. If Entrance exam result = Very bad and Grades = Good, then Student evaluation = Bad.
5. If Entrance exam result = Very bad and Grades = Excellent, then Student evaluation = Bad.
6. If Entrance exam result = Bad and Grades = Very bad, then Student evaluation = Bad.
7. If Entrance exam result = Bad and Grades = Bad, then Student evaluation = Bad.
8. If Entrance exam result = Bad and Grades = Average, then Student evaluation = Average.
9. If Entrance exam result = Bad and Grades = Good, then Student evaluation =Average.
10. If Entrance exam result = Bad and Grades = Excellent, then Student evaluation = Average. Etc.

To meet the new requirements of the university, only the rules number 3, 4 and 5 have to be modified. The terms on their right hand sides will be changed to *Very bad*. Again, it can be seen, that the use of the proposed algorithm saves time and effort of the decision-maker, who is required to modify just 3 rules instead of setting all 25 of them.

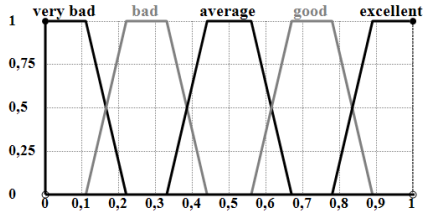


Figure 1: Linguistic variable for the evaluation

Acknowledgements

The research has been supported by GACR project with the registration number 14-02424S.

6 Conclusion

In the paper, the generalized partial goals method and its software implementation is described briefly. The main part of the paper deals with the situation when the evaluation model is being tuned up. If the decision-maker has to change one aggregation method for a more complicated one, the parameters for the new method will have to be set. Some methods however require quite a large number of parameters. In this paper, we have suggested methods that can make this task much easier. The parameters for the new aggregation method are derived from the parameters of the original one. The decision-maker then just modifies some of those parameters instead of setting all of them. This way, the change can be done in a much easier way, which has been demonstrated on the examples.

References

[1] D. Dubois and H. Prade. *Fundamentals of Fuzzy Sets. The Handbook of Fuzzy Sets Series.* Kluwer Academic Publishers, Boston-London-Dordrecht, 2000.

[2] M. Grabisch and C. Labreuche. A decade of application of the Choquet and Sugeno integrals in multicriteria decision making. *Annals of Operations Research 175*, pages 247–286, 2010.

[3] P. Holeček, J. Talašová, and I. Müller. Fuzzy methods of multiple-criteria evaluation and their software implementation. In V. Mago and N. Bhatia, editors, *Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies*, pages 388–411. IGI Global, 2012. ISBN: 978-1-61350-429-1, DOI: 10.4018/978-1-61350-429-1.ch021.

[4] O. Pavlačka and J. Talašová. The fuzzy weighted average operation in decision making models. *Proceedings of the 24th International Conference Mathematical Methods in Economics, 13th - 15th September 2006, Plzeň (Ed. L. Lukáš)*, pages 419–426, 2006.

[5] J. Talašová. *Fuzzy methods of multiple criteria evaluation and decision making (in Czech)*. Publishing House of Palacký University, Olomouc, 2003.

Decomposition of fiscal imbalance in EU member countries

Sarka Horakova[1], Robert Jahoda[2]

Abstract. The stability of the common currency requires coordination of fiscal policies within the EU through compliance with agreed rules in the Stability and Growth Pact (SGP). Member countries stipulate the obligation to keep the public deficit below 3 % of GDP and public debt below 60 % of GDP. Enforcement of SGP is, however, debatable and with regard to public finances of member countries in the last few years its effectiveness could be questioned. Agreed rules should lead to long-term balanced and sustainable public finances. It would be correct to assume that the respect of agreed rules would lead to long-term balanced budgets. In our previous research we established an indicator A1 which gives a level of non-compliance with the required rules. Surprisingly, it was found that despite the compliance with required criteria a large number of countries would be facing an increase in the level of debt. In this paper, we are focused on closer examination of that phenomenon by decomposition of the level of debt. By simulation modelling, we determine the impact of individual factors causing the raise of indebtedness.

Keywords: fiscal imbalance, Stability and Growth Pact, primary balance, public debt, European Monetary Union.

JEL Classification: C63, E62, H62, H63
AMS Classification: 90C15

1 Introduction

The suitability of the Stability and Growth Pact as a tool of fiscal coordination was contested since its adoption. One would suppose that after the reform from 2005 the modified provisions would regain its credibility. But the financial crisis erupted in 2007/2008 and transformed in to the debt crisis when several European countries 'public finance had proven to be unsustainable. At this point the Stability and Growth Pact failed as a tool to maintain the fiscal discipline in European Union. Despite the Maastricht criteria and the SGP, many states have failed to reduce the value of their debt to a sustainable level (Van Nieuwenhuyze 2013). Stoian and Alves (2012) also show in their study that the euro area does not follow a fiscal policy that would be compatible with the requirements of the SGP. One of the consequences of the debt crisis is that risk can be determined for each country separately, while bonds are denominated in the common currency. This contradicts the principles of the EMU, which were established in the 1990s (Buiter 2006). The economy of member countries is very heterogeneous and, furthermore, we cannot talk about fiscal federalism in the EU. The euro seems overvalued for southern European countries (including France) and underestimated for countries in Northern Europe, especially Germany (Coudert 2013; Duwicquet et al. 2013). Financial markets may decide to refuse a loan to some countries, simply because the operation would be too risky. Consequently, heavily indebted countries may find themselves close to default. The rules to coordinate fiscal policies were established in SGP. The value of the 60% debt/GDP ratio was determined based on the average of the members, and the 3% threshold for the deficit/GDP ratio is based on a simple dynamic model for public debt. According to De Grauwe (2003, p. 2), "it is well known that the 3% deficit norm will ensure that the 60% debt ratio can be kept constant provided the nominal growth of GDP happens to be 5%". It is necessary to respect this value to avoid constant growth of debt and the snowball effect.

Using a simple public debt dynamic model we separate the influence of compliance and non-compliance with the maximum level of the deficit rule relative to GDP. Our previous research surprisingly showed that most of EU countries would suffer the raise of the debt level despite the respect of agreed rules in period 1995-2012. In our paper, we will analyze factors causing this situation and we will examine if it is possible to separate the influence of individual factors.

Stoian and Alves (2012) use a similar model to calculate the level of government surplus that would be necessary in order to stabilize the public debt. Stoian (2011) believes that the economic crisis may have been instigated due to improper fiscal policy. He uses the dynamic model of debt to determine whether the conduct of fiscal policy is vulnerable or not. Afonso (2005) uses the model of a stationary test to determine the fiscal sustainabil-

[1] Masaryk University, Department of public economics, Lipová 41a, Brno 60200, sarka.horakova@econ.muni.cz.
[2] Masaryk University, Department of public economics, Lipová 41a, Brno 60200, jahoda@econ.muni.cz

ity of all the Eurozone member countries and, like most authors, he agrees that the implementation of fiscal policy is not sustainable. Using this model, Izák (2009) confirms that a government that runs budget deficits faces significantly higher costs of borrowing. Van den Noord (2010) attempts to quantify costs of the financial crisis and the impact on the sustainability of public finance.

2 Methodology

In our previous research (Horakova and Jahoda 2013), we constructed the simulation model to assess the impact of non-compliance with agreed rules on the level of debt, in other words in what extent the irresponsible fiscal policy affects the evolution of public debt. The root of the model (1) is based on the equation of time for public debt, which stipulates that the level of public debt in the current year (marked as B^t) is equal to the level of debt in the previous year and this year's change in debt.

We use two basic indices. The upper index *t* marks a year for which the value is calculated or simulated. The lower index S is used to describe the simulated variable or is missing in case when the variable is calculated from the available data (see the end of this section).

Ideally, the change in debt (equation 2) is equivalent to the level of public deficit in a given year (D^t). Afonso (2005) draws attention to the fact that the transfer of public deficit into debt is not necessarily time-consistent within one year. Therefore, we introduce and compute the variable SFA^t, which captures all the situations in which the growth of public debt in the reference year is different from the size of the public deficit.

$$B^t = B^{t-1} + \Delta B^t \tag{1}$$

$$SFA^t = \Delta B^t - D^t \tag{2}$$

The public deficit (in equation 3) can be expressed as the sum of the primary balance (PD^t) and the interest on public debt (IB^t).

$$D^t = PD^t + IB^t \tag{3}$$

We use a simplified relationship in the model (equations 4) to show that the amount of interest in a given year depends on the amount of the debt at the end of the previous year and the implicit interest rate (r^t) of the debt. Equation 5 calculates the annual real growth of the gross domestic product (y^t) from data on GDP at constant (2005) prices (GDP_*).

$$IB^t = r^t * B^{t-1} \tag{4}$$

$$y^t = \frac{GDP_*^t}{GDP_*^{t-1}} - 1 \tag{5}$$

Equation 6 provides a basic assumption on which the model stands, namely that the value of the public deficit in each year should be lower than 3% of GDP.

$$d_S^t \leq 3\% \tag{6}$$

We assume (see equation 7) that the simulated level of interest on public debt is due to the simulated level of public debt and the amount of the implicit interest rate.

$$IB_S^t = r^t * B_S^{t-1} \tag{7}$$

Equations 1 and 2 are used to describe the decomposition of the actual government debt; equations 8 and 9 then return to compose the simulated level of debt. Equation 8 therefore states that the simulated increase in public debt is equal to the simulated deficit and the "time discrepancies" that we have excluded in equation 2 from the other calculations. Finally, equation 9 describes the simulated amount of debt at the end of the budget period and is equal to the simulated value of the debt at the beginning of the period and the simulated growth of public debt in the given period.

$$\Delta B_S^t = D_S^t + SFA^t \tag{8}$$

$$B_S^t = B_S^{t-1} + \Delta B_S^t \tag{9}$$

We evaluate the quality of the fiscal policy of each country in the past 17 years, by comparing the amount of the real and the simulated level of public debt (equation 11). For reasons of comparability, we always use a GDP ratio.

$$B_S^{1995} = B^{1995} \tag{10}$$

$$A1_j = \frac{B_S^{2012} - B^{2012}}{GDP^{2012}} \tag{11}$$

The model is based on Eurostat data for the period 1995–2012. We are limited to the year 1995 because it is the first year for which the data are available in the Eurostat database. For all the Eurostat data, the ESA95 methodology was applied. Since Eurostat does not provide data on GDP at constant prices for the whole period 1995–2012, we used data from United Nation Data – GDP by Type of Expenditure at constant (2005) prices – National currency. UN Data uses the very same methodology for GDP calculation as Eurostat.

Our presented paper is based on the previous model and elaborates the individual components influencing the public debt. The share level of debt/GDP can be decomposed on initial level of debt/GDP in 1995, the share deficit/GDP and the share of SFA/GDP.

$$\frac{B^{1995+i}}{GDP^{1995+i}} = \frac{B^{1995}}{GDP^{1995+i}} + \frac{\sum_{j=1}^{i} D^j}{GDP^{1995+i}} + \frac{\sum_{j=1}^{i} SFA^j}{GDP^{1995+i}} \tag{12}$$

Decomposition of the amount of debt in year (1995+i) is shown in equation 12. It was created by combination of equations 1 and 2 and debt was transformed in to the relative form to GDP. The individual components from the right side of the equation 12 are shown in equations 13 and 14.

$$\frac{B^{1995}}{GDP^{1995+i}} = \frac{B^{1995}}{GDP^{1995}} - \frac{B^{1995} * (GDP^{1995+i} - GDP^{1995})}{GDP^{1995} * GDP^{1995+i}} \tag{13}$$

The proportion of the original amount of debt to GDP in 1995 in assessed year is shown in the equation 14. We are able to capture the effect of change in GDP on relative reduction to the initial level of debt.

$$\frac{\sum_{j=1}^{i} D^j}{GDP^{1995+i}} = \frac{\sum_{j=1}^{i} D_{SGP*}^j}{GDP^{1995+i}} + \frac{\sum_{j=1}^{i} D_{SGP\_E}^j}{GDP^{1995+i}} + \frac{\sum_{j=1}^{i} D_{SGP\_NC}^j}{GDP^{1995+i}} \tag{14}$$

Equation 14 shows the decomposition of the accumulated deficit on absolute compliance with the maximum level of deficit of 3% of GDP (D_{SGP*}^j) component, the part of the deficit above 3% of GDP in moments when, due to the economic downturn, country does not need comply with SGP ($D_{SGP\_E}^j$) and finally the part of the deficit, which is due to the failure of compliance with the rules of the SGP ($D_{SGP\_NC}^j$).

Monitored impact of the economy on the level of public debt is demonstrated in equation 13 and equation 14.

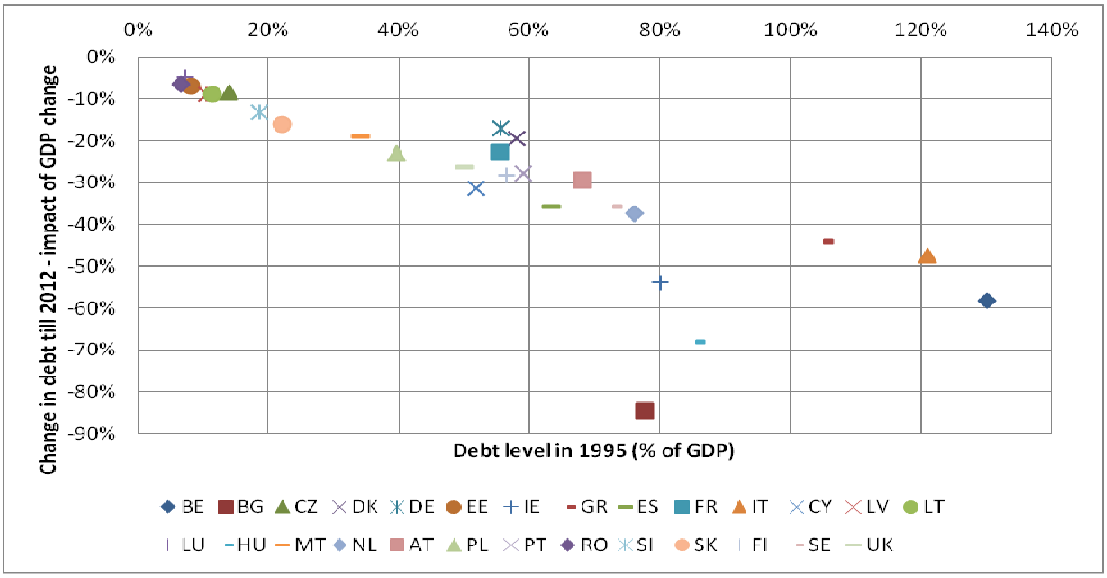
3 Results

In our paper we focused on level of indebtedness in EU countries. By its decomposition we have separated individual factors causing the raise in debt: the change in debt for the period 1995-2012, the A1 indicator (stands for irresponsible fiscal policy), Stock Flow Adjustments (SFA), Exceptions of SGP given by economic downturn, compliance with agreed rules (max deficit 3% of GDP) and the economic growth. Our previous paper (Horáková, Jahoda, 2013) was dedicated to the A1 indicator e.g. the impact of non-compliance with agreed rules. Present paper is focused on analysis of other components causing the raise of the debt level.

Figure 1 shows that the economic growth between 1995 and 2012 alone (the right side of equation 14) is strong enough to reduce the public debt level in all member countries. Figure shows a clear correlation between the

level of debt reduction and its initial level. Bulgaria and Hungary are exceptions, a significant change in price level is added to the real GDP growth (these two components influence the nominal GDP growth

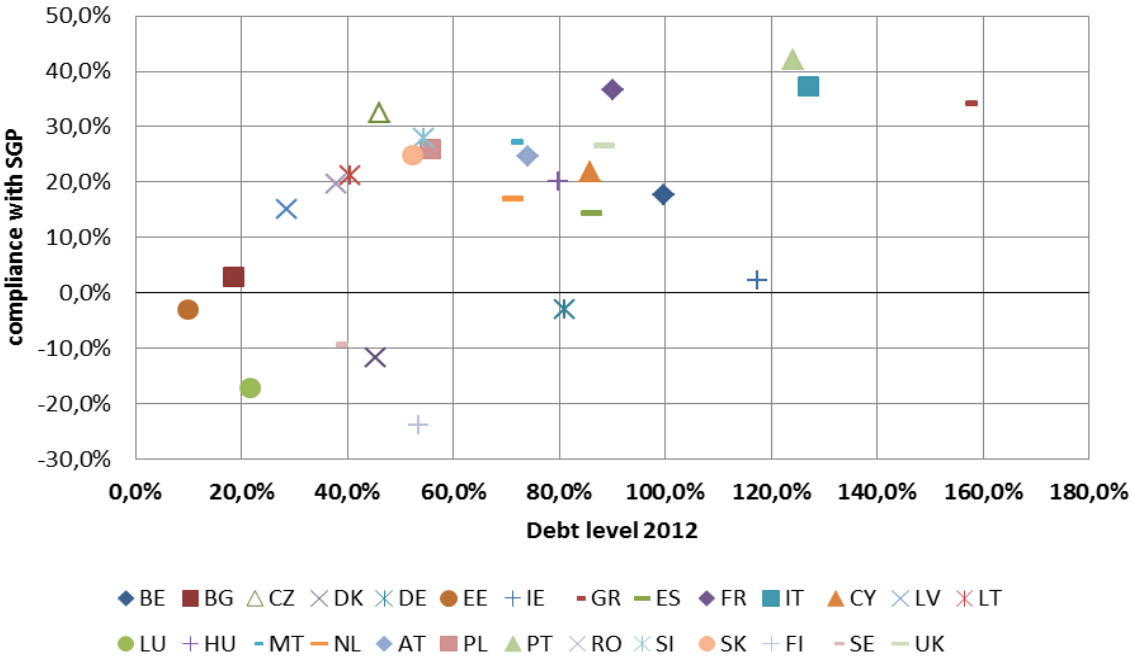
Figure 1: Change in debt till 2012 including GDP change impact



Source: own calculations

Figure number 2 depicts countries trying to comply with the SGP rules between 1995-2012 (see equation 12 and 14) which has ended with the lower debt level in 2012.

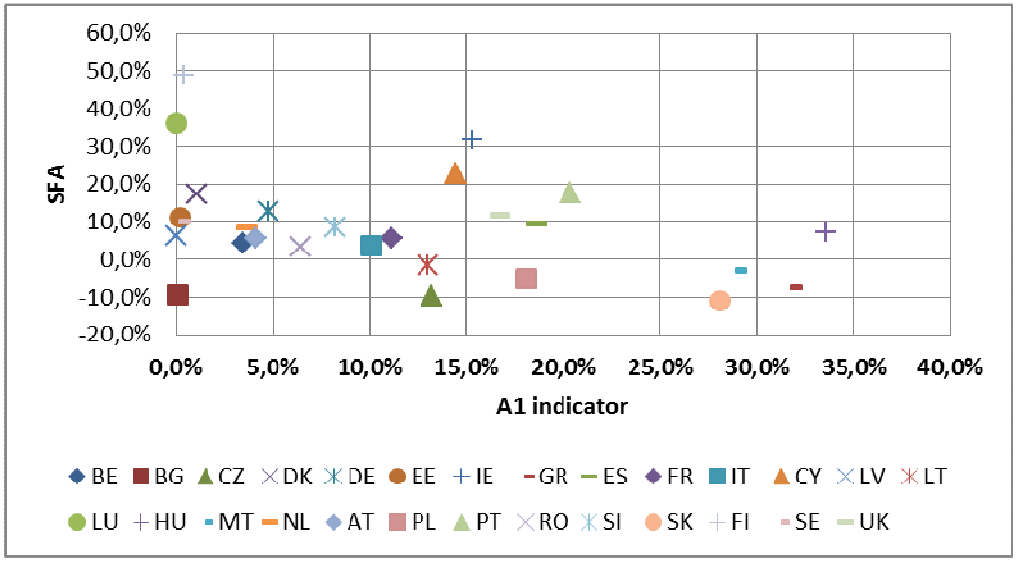
Figure 2: Simulated level of debt in 2012 under condition of compliance with SGP rules



Source: Own calculations

Figure 3 illustrates the relation between A1 indicator standing for influence of irresponsible fiscal policy and Stock flow adjustment.

Figure 3: Relation between irresponsible fiscal policy and SFA

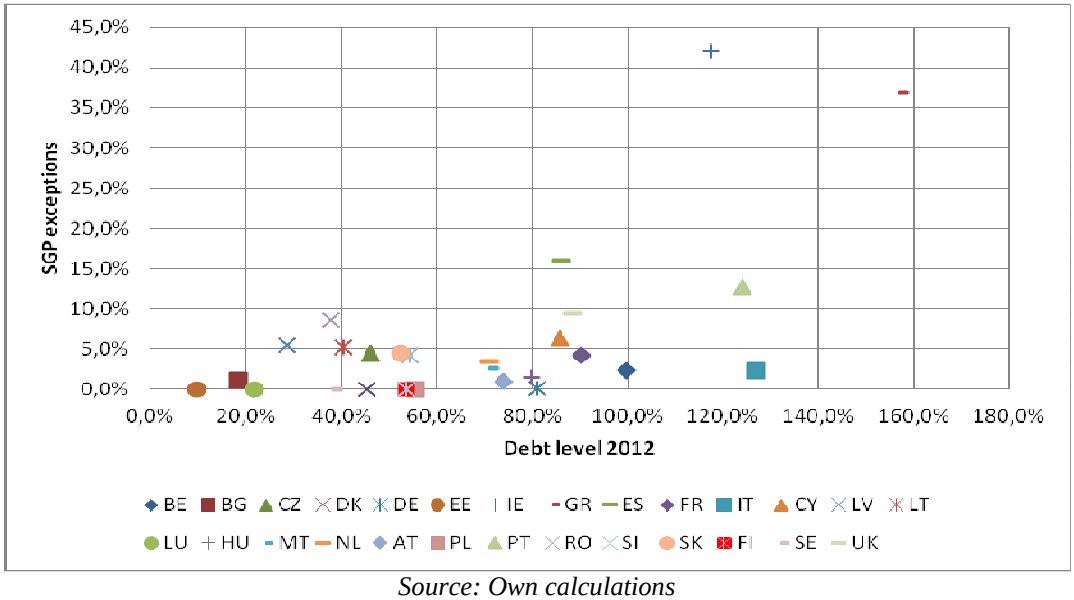


Source: Own calculations

The dependence is not very strong, but figure shows that there is a trade-off between compliance and the amount of accumulated SFA (Equation 13). Or it seems that countries that did not want to be accused of irresponsible fiscal policies tend to solve their fiscal imbalance by using SFA (creative accounting). As we stated in our previous research (Horakova and Jahoda, 2013) Eurostat has divided SFA into the following components: net acquisition of financial assets, debt adjustment effects and statistical discrepancies. The importance of SFA has been emphasized many times, because it can highlight data quality problems. It has been argued that since great attention is paid to the deficit under the current EU multilateral fiscal surveillance (EDP and Stability and Growth Pact); governments may have an incentive to underreport their deficits by reporting transactions under SFA. Accordingly, we can draw conclusion that the line between SFA and "creative accounting" is very thin and unclear.

As another factor explaining the raise of indebtedness we identified the exceptions from SGP allowing deficit higher than 3% GDP due to an economic downturn. Countries excusing their high deficits with unfavorable economic environment reached a significant level of indebtedness in 2012.

Figure 4: Impact of exceptions from SGP



Source: Own calculations

Therefore new rules do not accentuate overall deficit (+ exceptions) anymore but focuse on the level of structural deficits. With the new rules, there is no need to make excuses to adverse economic conditions, because they are

cyclically repeated. On the contrary, it is necessary to not breach the structural component and - 0.5% of GDP which is fairly strict.

4 Conclusion

The Stability and Growth Pact was supposed to guarantee the stability of common currency and fiscal discipline within the European Union. Even if its suitability was contested since its adoption the recent financial and economic crisis showed that this tool is not sufficiently bounding. Our paper is based on the hypothesis that the total level of debt could be decomposed on several components: the change in debt for the period1995-2012, the A1 indicator (stands for irresponsible fiscal policy), Stock Flow Adjustments (SFA), Exceptions of SGP given by economic downturn, compliance with agreed rules (max deficit 3% of GDP) and the economic growth. We showed that the nominal GDP growth should contribute to decrease of indebtedness but in reality it was not the case for EU countries (Figure 1). Our model showed that despite the respect of agreed rules (3% limits for the deficit) the vast majority of member states would recorded an increase of the debt level (figure 2). That is why public finances in some of them had shown to be unsustainable. We therefore analyze individual factors causing this phenomenon. Stock flow adjustment (Figure 3) caused the greatest raise in level of debt in Finland, Lithuania or Ireland e.g. states where the impact of irresponsible fiscal policy is the smallest. As we state earlier, SFA should be monitored closely since the line between SFA and "creative accounting" is very thin and unclear. The last component of the debt we examined was the economic downturn taken in to consideration by the SGP is particularly significant for Ireland, Greece, Portugal and Spain (Figure 4).

Acknowledgements

The authors are thankful to the Grant Agency of Masaryk University for the grant No. MUNI/A/0784/2013.

References

[1] AFONSO, A. (2005). Fiscal Sustainability: The Unpleasant European Case. Finanzarchiv. 61(1), pp. 19–44.
[2] BUITER, W. (2006). The "Sense and Nonsense of Maastricht" Revisited: What Have we Learnt about Stabilization in EMU? Journal of Common Market Studies. 44(4), pp. 687–710.
[3] BUITER, W. (2006). The "Sense and Nonsense of Maastricht" Revisited: What Have we Learnt about Stabilization in EMU? Journal of Common Market Studies. 44(4), pp. 687–710.
[4] COUDERT, V. (2013). On Currency Misalignments within the Euro Area. Review of International Economics. 21(1), pp. 35–48.
[5] DE GRAUWE, P. (2003). The Stability and Growth Pact in Need of Reform. Centre for European Policy Studies. Retrieved February 5, 2014, from :http//www.econ.kuleuven.ac.be/ew/academic/intecon/Degrauwe/PDG-papers/Discussion_papers/Stability%20and%20Growth%20Pact%20CEPS%202003.pdf
[6] DUWICQUET, V., MAZIER, J., SAADAOUI, J. (2013). Désajustement de change féderalisme budgétaire et redistribution: Comment s'ajuster en union monétaire. Revue de l'OFCE. 1(127), pp. 57–96.
[7] EUROSTAT (2013). The Main Factors Contributing to Changes in Government Debt Other than Government Deficits/Surpluses (Stock-Flow Adjustment). Retrieved February 5, 2014, from http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/STOCK_FLOW_2013_APR/EN/STOCK_FLOW_2013_APR-EN.PDF
[8] HORAKOVA, S., JAHODA, R. (2013) The SGP - faulty by design or made faulty by politicians? An assessment based on simulation model. In Proceedings of the 31st International Conference Mathematical Methods in Economics 2013. Jihlava : College of Polytechnics Jihlava, 2013. ISBN 978-80-87035-76-4, s. 279-284. 2013,
[9] IZÁK, V. (2009). Primary Balance, Public Debt and Fiscal Variables in Postsocialist Members of the European Union. Prague Economic Papers 2/2009.
[10] STOIAN, A. (2011). A Retrospective Approach on Assessing Fiscal Vulnerability: Empirical Evidence for Overindebted European Countries. Timisoara Journal of Economics. 4(3(15)), pp. 183–188.
[11] STOIAN, A., ALVES, R. (2012). Can EU High Indebted Countries Manage to Fulfill Fiscal Sustainability? Some Evidence from the Solvency Constraint. FEP working papers. Retrieved February 5, 2014, from http://wps.fep.up.pt/wps/wp464.pdf
[12] VAN DEN NOORD, P. (2010). Turning the Page? EU Fiscal Consolidation in the Wake of the Crisis. Empirica. 38(1), pp. 19–51.
[13] VAN NIEUWENHUYZE, C. (2013). Debt, Assets and Imbalances in the Euro Area. Revue de l'OFCE. 1(127), pp. 123–152. Retrieved February 5, 2014, from http://www.cairn.info/revue-de-l-ofce-2013-1-page-123.htm

Strong Nash general equilibrium

Milan Horniaček[1]

Abstract. We analyze an economy consisting of oligopolistic firms and consumers. Firms can produce several consumption goods and their production possibility sets need not be convex. Consumers make decisions on purchases of consumption goods and supply of labor services on the basis of utility maximization. Firms maximize real wealth of their shareholders. We identify and prove a sufficient condition for existence of a strong Nash general equilibrium among firms.

Keywords: general equilibrium, oligopoly, shareholders' real wealth, strong Nash equilibrium.

JEL Classification: C72, D43, D59

AMS Classification: 91A10, 91A80

1 Introduction

Most of the real world markets are characterized by market power on at least one side. Despite this, standard general equilibrium models (i.e., Arrow - Debreu - Hahn type of models; see, for example, [1] for description of these models) follow the paradigm of perfect competition. Thus, general equilibrium models that incorporate the interaction between imperfectly competitive firms are needed. So far, the paper by Dierker and Grodal [4] is the most comprehensive result in this direction.

In the present paper, we analyze an oligopolistic economy with a finite number of consumption goods and labor services. Consumes make decisions on purchases of consumption goods and on supply of labor services on the basis of utility maximization. Firms maximize shareholders' real wealth. We analyze a strong Nash general equilibrium (henceforth, SNGE) of this oligopolisitic economy. It is the application of Aumann's [2] concept of strong Nash equilibrium to general equilibrium.

2 Model

Throughout the paper, R denotes the set of real numbers. For $(b,c) \in R^n \times R^n$, $b \geq c$ means $b_k \geq c_k$ for each $k \in (1,...,n)$, $b > c$ means $b \geq c$ but $b \neq c$, $b >> c$ means $b_k > c_k$ for each $k \in \{1,...,n\}$. R^n_+ is the nonnegative orthant of R^n, i.e., $R^n_+ = \{b \in R^n \mid b \geq 0\}$, and Δ_n is the simplex in R^n, i.e., $\Delta_n = \left\{b \in R^n_+ \mid \sum_{k=1}^n b_k = 1\right\}$. We endow each finite dimensional real vector space with the Hausdorff topology compatible with the linear structure. For $(b,c) \in R^n \times R^n$, $[b,c]$ is the line segment joining b and c.

The analyzed economy E consists of the finite set of firms J with $\#(J) \geq 2$ and the finite set of consumers I with $\#(I) \geq 2$. A is the finite set of goods. $A = Q \cup L$, where Q is the non-empty set of consumption goods and L is the non-empty set of labor services.

We view E as the single period model of an economy. Thus, in E firms produce only consumption goods using hired labor services. We do not include produced intermediate goods in A. It is tantamount to assuming that each firm has initial stocks of produced intermediate goods that are parameters of the model and are fully used in production. We denote price of good $a \in Q$ by p_a, price paid for labor service $a \in L$ by firm $j \in J$ by p_{aj}, and assume that $p = \left((p_a)_{a \in Q}, \left((p_{aj})_{j \in J} \right)_{a \in A} \right) \in \Delta_{\#(Q)+\#(L)\times\#(J)}$.

Each $i \in I$ has endowment $\omega_i = (\omega_{ia})_{a \in A} \in R^{\#(A)}_+ > 0$. For each $i \in I$, $X_i \subset R^{\#(Q)}_+$ is consumer i's consumption set. We assume that it is convex and $(\omega_{ia})_{a \in Q} \in X_i$. The symbol ℓ_{ija} stands for the supply

[1]Comenius University in Bratislava, Faculty of Social and Economic Sciemces, Institute of Economics, milan.horniacek@fses.uniba.sk

of labor service $a \in L$ by consumer $i \in I$ to firm $j \in J$. Of course, for each $(i,a) \in I \times L$, $\ell_{ija} \geq 0$ for each $j \in J$ and $\sum_{j \in J} \ell_{ija} \leq \omega_{ia}$. (We assume that each $i \in I$ can provide all labor services specified by $(\omega_{ia})_{a \in L}$.) For each $(i,j) \in I \times J$, $\alpha_{ij} \in [0,1]$ is consumer i's ownership share in firm j; $\sum_{i \in I} \alpha_{ij} = 1$ for each $j \in J$.

For each $j \in J$, $Y_j \subset R^{\#(A)}$ is firm j's production possibility set. For each $y_j = (y_{ja})_{a \in A} \in Y_j$, $y_{ja} \geq 0$ is the output of consumer good a if $a \in Q$ and $-y_{ja} \geq 0$ is the used quantity of labor service a if $a \in L$.

Assumption 1. *For each $j \in J$ (i) $0 \in Y_j$, (ii) $y_{ja} \leq 0$ for each $a \in L$ and each $y_j \in Y_j$, (iii) $y_{ja} \geq 0$* for each $a \in Q$ and each $y_j \in Y_j$, *(iv) $y_{ja} = 0$ for each $a \in L$ implies $y_j = 0$, (v) Y_j is closed, (vi) if $y_j \in Y_j$ and $y'_j < y_j$, then $y'_j \in Y_j$, (vii) the boundary of Y_j does not contain an infinite portion that is a subset of the vector space with the dimension lower than $\#(A)$ and containing at least one coordinate axis.*

Part (i) of Assumption 1 implies that Y_j is non-empty. Part (ii) stems from the fact that labor services cannot be produced. Part (iii) stems from the fact that consumption goods are not used as inputs of production. Part (iv) states that it is impossible to produce something without using any labor service. Part (vi) is the assumption of free disposal. It implies that increases in labor inputs cannot decrease outputs. The following assumption expresses the fact that all consumption goods are differentiated, i.e., each of them can be produced by exactly one firm.

Assumption 2. *For each $a \in Q$ there exists $\iota(a) \in J$ and $y_{\iota(a)} \in Y_{\iota(a)}$ such that $y_{\iota(a)a} > 0$ and $y_{ka} = 0$ for each $k \in J \setminus \{\iota(a)\}$ and each $y_k \in Y_k$.*

The price of each consumption goods $a \in Q$ is set by the firm that is capable of producing it. Consumers with a positive endowment of a can sell part or all of it at the latter price.

The following assumption states that it is possible to produce simultaneously positive amount of each consumption good.

Assumption 3. *For fixed $\left((\omega_{ia})_{a \in L}\right)_{i \in I}$ there exists $(y_j)_{j \in J}$ such that $-\sum_{j \in J} y_{ja} \leq \sum_{i \in I} \omega_{ia}$ for each $a \in L$ and $y_{\iota(a)a} > 0$ for each $a \in Q$.*

Part (vii) of Assumption 1 implies that, for given fixed $\left((\omega_{ia})_{a \in L}\right)_{i \in I}$, the output of each consumption good is bounded from above. Thus, taking into account Assumption 2, for each $a \in Q$ there exists $b_a^{\max} > 0$ such that $y_{\iota(a)a} \leq b_a^{\max}$ for each $y_{\iota(a)} \in Y_{\iota(a)}$. This justifies the following simplifying assumption. It allows us to shorten the paper but the existence of an SNGE does not depend on it.

Assumption 4. *For each $i \in I$ and each $a \in Q$, each $x_i \in X_i$ satisfies $x_{ia} \leq b_a^{\max}$.*

For each $i \in I$ we set

$$\mathcal{L}_i = \left\{ \ell_i = \left((\ell_{ija})_{j \in J} \right)_{a \in L} \in R_+^{\#(J) \times \#(L)} \mid \sum_{j \in J} \ell_{ija} \leq \omega_{ia} \; \forall \, a \in A \right\}.$$

We do not view leisure as a consumption good. Nevertheless, each consumer has preferences not only over elements of his consumption set but also over supply of labor services to firms. His preferences over provision of labor services depend not only on the type and amount of the provided labor service but also on the identity of the firm to which he provides it. Thus, preferences of $i \in I$ are defined on the set $X_i \times \mathcal{L}_i$ and are represented by utility function $u_i : X_i \times \mathcal{L}_i \to R$.

Assumption 5. *For each $i \in I$ u_i (i) is continuous, (ii) it represents locally non-satiated preferences, (iii) it is non-decreasing in the consumed amount of each consumption good, (iv) it is strictly concave.*

For each $j \in J$ let

$$\Pi_j = \left\{ r \in R_+ \mid \exists \; p \in \Delta_{\#(Q)+\#(L)\times\#(J)} \; \& \; y_j \in Y_j \text{ such that } r = py_j \right\}$$

and set $\Pi = \prod_{j \in J} \Pi_j$. Thus, Π_j is the set of feasible non-negative profits of firm j (i.e., the set of feasible sums of dividends) computed at price vectors normalized to the simplex. For each $i \in I$, $d_i : \Delta_{\#(Q)+\#(L)\times\#(J)} \times \Pi \to X_i \times \mathcal{L}_i$ is consumer i's Marshallian demand - labor supply function. It assigns to each price vector $p \in \Delta_{\#(Q)+\#(L)\times\#(J)}$ and each vector of sums of dividends i's demand for consumption goods and his supplies of labor services to individual firms. It is defined by

$$d_i(p,\pi) = \arg\max \left\{ \begin{array}{c} u_i(x_i,\ell_i) \mid x_i \in X_i, \ell_i \in \mathcal{L}_i, \\ \sum_{a \in Q} p_a x_{ia} \leq \\ \sum_{a \in Q} p_a \omega_{ia} + \sum_{a \in L} \sum_{j \in J} p_{aj} \ell_{ija} + \sum_{j \in J} \alpha_{ij} \pi_j \end{array} \right\}. \tag{1}$$

Part (iv) of Assumption 5 (together with the assumption that X_i is convex) ensures that the maximization program in (1) has the unique solution. Thus, d_i is a function. It follows from the theorem of the maximum that d_i is a continuous function.

In the case of excess demand for some consumption good rationing using a fixed scheme continuous in original demands (e.g. proportional to demands) is used. The existence of an SNGE does not depend the applied continuous rationing scheme.

We denote by G the strategic form non-cooperative game modelling strategic relationships in E. Since consumers do not behave strategically, only firms in J are players in G. We restrict attention to pure strategies in G. S_j is the set of pure strategies of $j \in J$ in G and we set $S = \prod_{j\in J} S_j$. Let

$$L_j = \{a \in L \mid \exists\, y_j \in Y_j \text{ such that } y_{ja} < 0\}.$$

For each $j \in J$, each $s_j \in S_j$ specifies planned sum of dividends $\pi_j \in \Pi_j$, as well as price vector $p^{(j)} \in (\#(J))^{-1} \Delta_{\#(Q)+\#(L)\times\#(J)}$ satisfying $p_a^{(j)} = 0$ if either $a \in Q$ and $\iota(a) \neq j$ or $a \in L\backslash L_j$. Note that S is compact. The outcome of $s \in S$ is the collection $(y_j(s))_{j\in J} \in \prod_{j\in J} Y_j$ and $\pi(s) \in \Pi$ computed as follows. In the first step, we set $p(s) = \sum_{j\in J} p^{(j)}$ and $y_{ja}(s) = -\sum_{i\in I} \ell_{ija}$ for each $a \in L$. In the second step, we compute $d_i(p(s), \pi(s))$ for each $i \in I$. In the third step, for each $j \in J$ we compute

$$y'_j(s) = \arg\max \left\{ \begin{array}{c} k \in [0,1] \mid y'_j \in Y_j,\ y'_{ja} = y_{ja}(s)\ \forall a \in L, \\ y'_{ja} = k \max\left\{\sum_{i\in I} (d_{ia}(p(s), \pi(s)) - \omega_{ia}), 0\right\} \\ \forall a \in Q \text{ with } \iota(a) = j, \\ y'_{ja} = 0\ \forall a \in Q \text{ with } \iota(a) \neq j \end{array} \right\}. \tag{2}$$

With respect to continuity of functions d_i, part (vii) of Assumption 1, and Assumption 4, function $y'_j : S \to Y_j$ defined by 2 is continuous. In the fourth step, if for each $j \in J$ $y'_j(s)$ gives profit no lower than π_j, we set $y_j(s) = y'_j(s)$ and $\pi_j(s) = \pi_j$ for each $j \in J$. Otherwise, for each $j \in J$ we replace π_j by π'_j equal to the minimum from π_j and profit given by $y'_j(s)$ and return to step 1. We obtain $(y_j(s))_{j\in J}$ in a finite number of iterations or as a limit. If $(y_j(s))_{j\in J}$ does not fully cover demands, rationing is used to obtain consumption vectors generated by s.

Denote by $(\widehat{x}_i(s))_{i\in I}$ the collection of consumers' consumption vectors and by $(m_i(s))_{i\in I}$ collection of their incomes generated by s. Taking into account the above observations and continuity of the rationing scheme, $(\widehat{x}_i(s))_{i\in I}$ and $(m_i(s))_{i\in I}$ are continuous in s. For each $C \in 2^J \backslash \{\varnothing\}$ let

$$I_C = \left\{ i \in I \mid \sum_{j\in C} \alpha_{ij} > 0 \right\}.$$

.

Definition 1 *A strategy profile $s^* \in S$ is an SNGE of G if (a) $\sum_{j\in J} y_{ja}(s) + \sum_{i\in I} \omega_{ia} = \sum_{i\in I} \widehat{x}_{ia}(s)$ for each $a \in Q$ and (b) there do not exist coalition $C \in 2^J \backslash \{\varnothing\}$, $s_C \in \prod_{j\in J} S_j$, and $z \in R^{\#(Q)}$ with $z >> 0$ such that*

$$p\left(\left(s_C, s^*_{-C}\right)\right)\left(\sum_{i\in I_C} \widehat{x}_i(s^*) + z\right) \leq \sum_{i\in I_C} m_i\left(\left(s_C, s^*_{-C}\right)\right). \tag{3}$$

Part (ii) of Definition 1 corresponds to the definition of real wealth maximization on p. 266 in [4] for coalition C. Part (i) states that s^* should lead to an equilibrium in the market for consumption goods. (Note that by definition of $(y_j(s))_{j\in J}$ it leads to an equilibrium in the labor market.)

3 Existence of an SNGE

Proposition 1 *Game G associated with economy E has an SNGE.*

Proof. (Due to space limitations we only outline the proof.) Take arbitrary $\left(y^*_j\right)_{j\in J} \in \prod_{j\in J} Y_j$ such that $y^*_{\iota(a)a} > 0$ for each $a \in Q$, $-\sum_{j\in J} y^*_{ja} = \sum_{i\in I} \omega_{ia}$ for each $a \in L$, and there does not exist $(y_j)_{j\in J} \in \prod_{j\in J} Y_j$ with the property that $-\sum_{j\in J} y_{ja} \leq \sum_{i\in I} \omega_{ia}$ for each $a \in L$ and $\left(\sum_{j\in J} y_{ja}\right)_{a\in Q}$ weakly Pareto dominates $\left(\sum_{j\in J} y^*_{ja}\right)_{a\in Q}$. Let

$$\Psi = \left\{ \begin{array}{c} \psi = \left((x_{ia})_{a\in Q}, \left((\ell_{ija})_{j\in J}\right)_{a\in A}\right)_{i\in I} \in R_+^{\#(I)\times(\#(Q)+\#(L)\times\#(J))} \mid \\ \sum_{i\in I} x_{ia} = y^*_{\iota(a)a} + \sum_{i\in I} \omega_{ia}\ \forall a \in Q, \\ \sum_{j\in J} \ell_{ija} = \omega_{ia}\ \forall a \in L \end{array} \right\}. \tag{4}$$

Ψ is a non-empty compact convex subset of R^m, where

$$\begin{aligned} 4 &\leq m = \#(I) \times (\#(Q) + \#(L) \times \#(J)) - \#(Q) - \#(L) \\ &< \#(I) \times (\#(Q) + \#(L) \times \#(J)). \end{aligned}$$

Denote by Ψ^+ the vector space R^m containing Ψ. For each

$$(p, \psi) \in \Delta_{\#(Q)+\#(L)\times\#(J)} \times \Psi$$

let $\widetilde{\pi}(p, \psi)$ be the vector of firms' profits at price vector p, collection of output vectors $\left(y_j^*\right)_{j\in J}$, and labor supplies given by ψ. Set $\pi_j(p, \psi) = \max\{\widetilde{\pi}_j(p, \psi), 0\}$ for each $j \in J$. Let $\rho(p, \psi)$ be the Euclidean distance between ψ and
$(d_i(p, \pi(p, \psi)))_{i\in I}$. Define function $\xi : \Delta_{\#(Q)+\#(L)\times\#(J)} \times \Psi \to R_+$ as follows. Consider, for fixed $s \in S$ and $C \in 2^J \setminus \{\varnothing\}$, (3) with $\widehat{x}_i(s^*)$ replaced by i's consumption vector determined by ψ, s^* replaced by the strategy profile generating p, $\left(y_j^*\right)_{j\in J}$, and sums of dividends $\pi(p, \psi)$, and the condition $z >> 0$ replaced by $z \in R^{\#(Q)}$. For each z satisfying (3) with the modifications described above let

$$z_{\min}(s, C) = \max\{\min\{z_a \mid a \in Q\}, 0\}$$

and $\widetilde{\xi}(s, C)$ be the maximum of $z_{\min}$ over all z satisfying (3) with the modifications described above. Then

$$\xi(p, \psi) = \max\left\{\max\left\{\widetilde{\xi}(s, C) \mid s \in S\right\} \mid C \in 2^J \setminus \{\varnothing\}\right\}.$$

Let $\lambda(p, \psi)$ be the ratio of $\rho(p, \psi) + \xi(p, \psi)$ to the maximum value of this sum over $\Delta_{\#(Q)+\#(L)\times\#(J)} \times \Psi$. Denote by $\theta(\psi)$ the vector of consumed quantities of consumption goods and labor supplies of all consumers determined by ψ and by $\beta(p, \psi)$ the projection of $(1 + \lambda(p, \psi))\theta(\psi)$ on Ψ^+. Note that if ψ is a boundary point of Ψ then $[\psi, \beta(p, \psi)] \cap \Psi = \{\psi\}$. Take continuous functions $\varepsilon : \Delta_{\#(Q)+\#(L)\times\#(J)} \times \Psi \to [0, 1]$ and $h : \Delta_{\#(Q)+\#(L)\times\#(J)} \times \Psi \to \Psi^+$ such that $\lambda(p, \psi) > 0$ implies $\epsilon(p, \psi) > 0$ and $\beta(p, \psi) + \varepsilon(p, \psi) h(p, \psi) \neq \psi$ and if ψ is a boundary point of Ψ then $[\psi, \beta(p, \psi) + \varepsilon(p, \psi) h(p, \psi)] \cap \Psi = \{\psi\}$.

We define function $f : \Delta_{\#(Q)+\#(L)\times\#(J)} \times \Psi \to \Delta_{\#(Q)+\#(L)\times\#(J)} \times \Psi^+$ as follows. For each (p, ψ) from its domain $f(p, \psi) = (p, \phi)$, where

$$\phi = (1 - \lambda(p, \psi))\psi + \lambda(p, \psi)(\beta(p, \psi) + \varepsilon(p, \psi) h(p, \psi)).$$

Function f satisfies all assumptions of Theorem 2 in [3], p. 286. Therefore, by the latter theorem, it has a fixed point (p^*, ψ^*). Strategy profile s^* that prescribes prices given by p^*, outputs of consumption goods and use of labor services given by $\left(y_j^*\right)_{j\in J}$, and sums of dividends given by $\pi(p^*, \psi^*)$is an SNGE of G. ■

Acknowledgement 1 *The research reported in the present paper was financially supported by the grant VEGA 1/0181/12.*

References

[1] Arrow, K. J., and Hahn, F. H.: *General Competitive Analysis,* Holden and Day, San Francisco, CA, 1971.

[2] Aumann, R.: Acceptable points in general cooperative n-person games, *Annals of Mathematical Studies* **40** (1959), 287–324.

[3] Browder, F. E.: A new generalization of the Schauder fixed point theorem, *Mathematische Annalen* **174** (1967), 285–290.

[4] Dierker, E., and Grodal, B.: The price normalization problem in imperfect competition and the objective of the firm, *Economic Theory* **14** (1999), 257–284.

Random Walk Process and Its Invertibility

Richard Horský[1]

Abstract. The assumption of the stationarity of the stochastic process ensures its convergence. The spectral analysis of the lag operator shows that if we take the random walk process, i.e. the nonstationary process the terms of which are generated by the cummulation of random shocks, we obtain a divergent process in the space of the bounded sequences. However if we consider the same process in the space of square summable sequences we get an invertible process. Unfortunately, the stationary processes are not contained in this space.
This trouble can be removed when we consider the centered processes stationary up to certain time far in the past then having terms the variances of which tend to zero. It does not conflict with real computations since the data are always finite. The autoregressive operator of the random walk process is the difference operator. Its inverse in the Hilbert space of all square summable sequences $\ell_2(L_2)$ is unbounded operator whose domain is everywhere dense in $\ell_2(L_2)$. It is a source of instability. It seems to be suitable to use symmetrization. Thus we get the symmetric operator the inverse of which is self-adjoint and therefore closed operator.

Keywords: Random walk process, difference operator, square summable sequences, self-adjoint operator, closed operator.
JEL Classification: C220
AMS Classification: 91B70

1 Stochastic process in Hilbert coordinate space

Stochastic process is a mapping $\mathbf{X}:\Im\to M(\Omega,\pi)$, where $\Im$ is a time domain and $M(\Omega,\pi)$ is a suitable space of functions or rather random variables defined on some measurable space Ω equipped with the probability measure π. It seems to be appropriate to choose $M(\Omega,\pi)=L_2(\Omega,\pi)$, i.e. the space of all functions defined almost everywhere on the space Ω, square integrable over the space Ω with respect to the measure π. This space is Hilbert space with the norm derived from the scalar product $E(XY)=\int_\Omega XYd\pi$. The mean and variance of any random variable $X\in L_2(\Omega,\pi)$ are finite numbers, particularly the mean $\mu=EX=\int_\Omega Xd\pi$ and the variance $\sigma^2=DX=EX^2-\mu^2$. The time domain can be taken in different ways; here it will be the set of all integers. It means that the values of a stochastic process are random variables $X_t, t=0,\pm1,\pm2,\ldots$ We will write $\mathbf{X}=(X_t)$ to emphasize that the stochastic process is intended to be a sequence. An example of a stochastic sequence is the white noise process $\mathcal{E}=(\varepsilon_t)$ with mean $E(\varepsilon_t)=0$, variance $D(\varepsilon_t)=E(\varepsilon_t^2)=\sigma_\varepsilon^2>0$ and autocovariance $\gamma_k=\mathrm{cov}(\varepsilon_t,\varepsilon_{t-k})=E(\varepsilon_t\varepsilon_{t-k})$ for any integer k. It is obviously an orthogonal system in the Hilbert space $L_2(\Omega,\pi)$, which can be transformed to be orthonormal system, not necessary complete.

The stochastic processes form different sets. The space of all sequences the elements of which are random variables with finite mean and variance, is too large for our purposes. Even if it is possible to establish a convergence structure in this space it does not the structure of normed linear space which we prefer. It seems to be natural to take the space of all bounded sequences. This space contains all stationary linear processes and it may be reason for this choice. However, as we explain later, another choice of the space of stochastic processes seems to be suitable. We have in mind the space of all sequences square summable. It is in fact the Hilbert coordinate space ℓ_2 of all sequences, here their elements are not (only) numbers but the functions (random variables) from $L_2(\Omega,\pi)$.

1.1 General linear process and its stationarity

The notion of stationary process is well-known (see [3], p.285). The property of stationarity ensures a stochastic stability of the stochastic process. The white noise process is the typical example of the stationary process. The white noise $\mathcal{E}=(\varepsilon_t)$ is a base for definition of the general linear process.

[1] University/of Economics, department of mathematics, Ekonomická 957, 148 00 Praha 4, rhorsky15@seznam.cz.

The *general linear process* is the process of the form

$$X_t = \mu + \sum_{n=0}^{\infty} \psi_n \varepsilon_{t-n}, \tag{1}$$

where μ is a real constant and (ψ_n) is a sequence of real numbers called *weights of the process*, $\psi_0 = 1$. It is clear that μ is the common mean of all X_t. We can set $Y_t = X_t - \mu$, hence $EY_t = 0$. This is a reason for the assumption that $\mu = 0$ in (1). The stochastic process with $\mu = 0$ is called the *centered process.* The series in (1) is the Fourier series of the element X_t with respect to the orthogonal system $\mathcal{E} = (\varepsilon_t)$ in $L_2(\Omega, \pi)$. The convergence of this series is equivalent to the condition

$$\sum_{n=0}^{\infty} \psi_n^2 < \infty. \tag{2}$$

This is in considerably more details described in [2], p.286 or in more general approach in [4], p.116. Due to (2) and (1) there is an isometric isomorphism between closed linear span of a particular white noise $\mathcal{E} = (\varepsilon_t)$ in $L_2(\Omega, \pi)$ and the Hilbert coordinate space ℓ_2, the elements of which are all numerical sequences (ψ_n) satisfying (2). It follows for instance that the scalar products are equal: the autocovariance function of the process (1) in terms of its weights

$$\gamma_k = E(X_t X_{t-k}) = E\left(\sum_{n=0}^{\infty} \psi_n \varepsilon_{t-n}\right)\left(\sum_{n=0}^{\infty} \psi_n \varepsilon_{t-k-n}\right) = \sigma_\varepsilon^2 \sum_{n=0}^{\infty} \psi_n \psi_{n+k}. \tag{3}$$

With respect to the well-known Schwartz inequality and the fact the l_2 – norm of (ψ_{n+k}) is at most equal to the l_2 – norm of (ψ_n) we obtain

$$|\gamma_k| = \sigma_\varepsilon^2 \left|\sum_{n=0}^{\infty} \psi_n \psi_{n+k}\right| \le \sigma_\varepsilon^2 \sum_{n=0}^{\infty} \psi_n^2 = \gamma_0. \tag{4}$$

The autocovariance function γ_k is a nonincreasing function on the set of nonnegative integers. Its upper bound is the variance of (1) $\gamma_0 = EX_t^2$ (independent on t). Thus if (2) holds, or equivalently (1) is convergent, then (4) implies that (1) is stationary. The opposite implication is obvious.

The general linear process (1) is thus uniquely determined by the sequence of the weights (ψ_n), $\psi_0 = 1$. This sequence is sometimes denoted as *the lag structure of the process* (1). A sufficient condition for the process (1) to be stationary is that its lag structure is absolutely convergent, i.e.

$$\sum_{n=0}^{\infty} |\psi_n| < \infty. \tag{5}$$

The condition (5) is not necessary for the process (1) to be stationary as the example of the sequence $\psi_n = \frac{1}{n}$ shows.

Example 1. If we set $\psi_n = \lambda^n$ we obtain so called *geometric lag*. It is obvious that the process (1) is stationary if and only if $|\lambda| < 1$. Then we say that the general linear process has a geometric lag structure.

Example 2. If $\psi_n = 0$ for any $n > q$ for some positive integer q, i.e. the lag structure has only finite number of nonzero elements, then the process (1) is obviously stationary.

1.2 Stationary process in $\ell_p = \ell_p(L_2)$ spaces

We will regard a stochastic process as a sequence of the square integrable functions (random variables)

$$\mathbf{X}= (X_t, X_{t-1}, X_{t-2}, \cdots) \tag{6}$$

for any $t \in \Im$. This assumption allows us to employ the well-known ℓ_p spaces, here its elements are of type (6). The set of all stationary processes is a subset of the space of all bounded processes $\ell_\infty = \ell_\infty(L_2) = \{\mathbf{X} = (X_{t-n}): X_{t-n} \in L_2(\Omega, \pi),\ \sup_{n\in\mathbb{N}_0}\|X_{t-n}\|_2 < \infty\}$. The norm $\|X_{t-n}\|_2$ is that in $L_2(\Omega, \pi)$ and for generally not centered stationary process $\mathbf{X} = (X_{t-n})$ with the mean μ and the variance σ^2 is given by

$$\|X_{t-n}\|_2^2 = EX_{t-n}^2 = \sigma^2 + \mu^2. \tag{7}$$

The norm of $\mathbf{X} \in \ell_\infty = \ell_\infty(L_2)$ is defined as $\|\mathbf{X}\|_\infty = \sup_{n\in\mathbb{N}_0}\|X_{t-n}\|_2$ for any $\mathbf{X} = (X_{t-n}) \in \ell_\infty$. A stationary process $\mathbf{X} = (X_{t-n})$ has the norm $\|\mathbf{X}\|_\infty = \sqrt{\sigma^2 + \mu^2}$. If a random variable $X \in L_2(\Omega, \pi)$ is a constant function, the stochastic process $\mathbf{X} = (X)$ is stationary with the norm $\|\mathbf{X}\|_\infty = |\mu|$. Our space $\ell_\infty = \ell_\infty(L_2)$ contains the bounded numerical sequences because they can be interpreted as the processes of constant random variables. If such a sequence is not constant the process is not obviously stationary.

Next we can consider the spaces $\ell_p = \ell_p(L_2) = \{\mathbf{X} = (X_{t-n}): X_{t-n} \in L_2(\Omega, \pi),\ (\sum_{n=0}^\infty \|X_{t-n}\|_2^p) < \infty\}$, $1 \le p < \infty$. The norm of the process $\mathbf{X} = (X_{t-n}) \in \ell_p = \ell_p(L_2)$ is defined similar to the classical case:

$\|\mathbf{X}\|_p = (\sum_{n=0}^\infty \|X_{t-n}\|_2^p)^{1/p}$. Unfortunately, the stationary processes are not contained in these spaces. It is caused by the fact that the necessary condition for the series $\sum_{n=0}^\infty \|X_{t-n}\|_2^p$ to be convergent is that

$$\lim_{n\to\infty}\|X_{t-n}\|_2 = 0 \tag{8}$$

for any fixed $t \in T$ and it is in contradiction with (7). For instance the typical stationary process, a white noise $\mathcal{E} = (\varepsilon_{t-n})$, satisfies $\|\mathcal{E}\|_p = (\sum_{n=0}^\infty \sigma_\varepsilon^p)^{1/p} = \infty$. This lack can be solved in the following way. We adopt another concept of stationarity. We allow the process to be stationary as far to the past as one wishes then the condition (8) has to start fulfilling. We can suppose only the centered processes as was mentioned. Then we are forced to relax the requirement of the constant variance in stationary process and with respect to (7) and (8) the variance of distant members of the process will converge to zero. We can call such process as the *stationary process up to the level N*. It means the process is stationary in classical sense only in the first *N* lags. For instance, in the case of the white noise up to the level *N*, $\mathcal{E} = (\varepsilon_{t-n})$, we have $\|\mathcal{E}\|_p^p = N{\sigma_\varepsilon}^p + \sum_{n=N+1}^\infty \sigma_{\varepsilon_{t-n}}^p$, where $\sum_{n=N+1}^\infty \sigma_{\varepsilon_{t-n}}^p < \infty$. The assumption of the independency of the members in $\mathcal{E} = (\varepsilon_{t-n})$ can be preserved. This access may be in accord to a real situation because the data are always finite and the assumption of constant mean or variance in any lag may seem to be rather artificial.

Among the spaces $\ell_p = \ell_p(L_2)$ there is unique Hilbert space. It is $\ell_2 = \ell_2(L_2)$. Its elements are all processes of type (6) satisfying the condition $\sum_{n=0}^\infty EX_{t-n}^2 < \infty$. It is equipped by the inner product $\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{n=0}^\infty E(X_{t-n}Y_{t-n})$. The norm derived from this inner product is $\|\mathbf{X}\|_2 = (\sum_{n=0}^\infty EX_{t-n}^2)^{1/2}$ and the space $\ell_2 = \ell_2(L_2)$ is complete with respect to this norm. This space will be employed in section 3.

2 The lag operator in $\ell_p = \ell_p(L_2)$ spaces

The theory of the stochastic processes is not conceivable without the introduction of the lag operator. This operator is extensively discussed in [3], p. 287. Here we define the *lag operator* as a mapping

$$B: \ell_p(L_2) \to \ell_p(L_2), \qquad B(X_{t-n}) = B(X_{t-1-n}).$$

2.1 The convergence structures

First, we take a process in the form (6) within $\ell_\infty = \ell_\infty(L_2)$. In this case we have $\|B\mathbf{X}\|_\infty \le \|\mathbf{X}\|_\infty$ for any process (6) and thus the lag operator *B* is bounded operator on $\ell_\infty = \ell_\infty(L_2)$. There is $\|B\| = \sup_{\|\mathbf{X}\|_\infty \le 1}\|B\mathbf{X}\|_\infty = 1$, since $B\mathbf{E} = \mathbf{E}$, where $\mathbf{E} = (1,1,1,\ldots)$, $\|\mathbf{E}\|_\infty = 1$. It holds $\|B^n\| = 1$ for any non-negative integer *n*. It follows that the space $[\ell_\infty]$ of all bounded operators contains the polynomials in *B*.

We can express any general linear process (1) using the lag operator in the form

$$\mathbf{X} = \psi(B)\mathcal{E}\,, \tag{12}$$

where

$$\psi(B) = \sum_{n=0}^{\infty} \psi_n B^n. \tag{13}$$

If (5) holds the formal power series in (13) is convergent in the norm of $[\ell_\infty]$ since

$$\|\psi(B)\| \le \sum_{n=0}^{\infty} |\psi_n| \|B^n\| = \sum_{n=0}^{\infty} |\psi_n|. \tag{14}$$

The operator (13) is called the *linear filter*. It transforms a white noise process to a general linear process.

Example 1. The linear filter is called the *geometric lag operator* if the sequence of its weights is of the form $\psi_n = \lambda^n$. The condition (5) is satisfied if and only if $|\lambda| < 1$. Let us notice that if we set $\lambda = 1$ we obtain divergent series (in the norm of the space $[\ell_\infty]$) $\sum_{n=0}^{\infty} B^n$. The partial sums of this series defines a divergent process called the *random walk process.*

Example 2. If there is a positive integer q such that $\psi_n = 0$ for any $n > q$, then (13) is of the form $\psi(B) = \sum_{n=0}^{q} \psi_n B^n$ and is called the *polynomial lag operator (of the degree q).*

Example 3. We can get a linear filter with infinite many nonzero weights which is not geometric (in fact it is a product of a finite number of geometric lag operators). Let us consider a stochastic linear difference equation of the order *p*

$$\Phi(B)\mathbf{X} = \mathcal{E}\,, \tag{15}$$

where $\Phi(B)$ is a polynomial lag operator of the degree *p*. Its solution is an *autoregressive process* and the $\Phi(B)$ is called the *autoregressive operator of the degree p*. The transformation (15) to the form (12) is possible iff all roots of the polynomial $\Phi(z)$ in the complex variable *z* lie outside the unit circle in the complex plane (see [2], p. 27). It can be also explained by the help of the functional calculus (see [5], p. 272).

Second, we consider the operator *B* on $\ell_p = \ell_p(L_2)$ where $1 \le p < \infty$. There is the basic difference (see 1.2): these spaces do not contain the stationary processes, e.g. $\mathbf{E} = (1,1,1,\dots) \notin \ell_p = \ell_p(L_2), 1 \le p < \infty$. To show that $\|B\| = \sup_{\|\mathbf{X}\|_p \le 1} \|B\mathbf{X}\|_p = 1$ we take the process $\mathbf{E_2} = (0,1,0,0,\dots)$.

The lag operator $B \in [\ell_p], 1 \le p \le \infty$ may be expressed in the matrix form

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{16}$$

The matrix representation (16) of the lag operator *B* is not necessary but it is suitable namely with its correspondence to the finite dimensional restrictions of a model of stochastic process.

2.2 Spectral properties of the lag operator

The spectral properties of the lag operator (16) in the spaces $\ell_p = \ell_p(L_2)$, $1 \le p \le \infty$ are described in [3], p. 290. Let us recall that in any case the spectrum of (16) is closed unit circle in the complex plane. Its structure is nevertheless different for $p = \infty$ and $p < \infty$. Firstly the sequence $\mathbf{E_\lambda} = (1, \lambda, \lambda^2, \lambda^3, \dots)$ is for any $|\lambda| < 1$ an eigenvector of (16). The boundary of the unit circle is also in the point spectrum of (16) for $p = \infty$. For instance if $\lambda = 1$ the corresponding eigenvector is the process $\mathbf{E} = (1,1,1,\dots)$. However if $p < \infty$ and $|\lambda| = 1$ the operator

$$(\lambda I - B)^{-1} \tag{17}$$

exists but it is not bounded and its domain is dense in the space $\ell_p = \ell_p(L_2)$, not equal to this space. The density of the space $\mathcal{R}(\lambda I - B)$ in $\ell_p = \ell_p(L_2)$ for $1 \le p < \infty$ follows from the fact that $\mathcal{R}(\lambda I - B)$ contains all the

sequences with a finite number of nonzero elements. However it cannot be $\mathcal{R}(\lambda I - B) = \ell_p = \ell_p(L_2)$, otherwise (17) would have to be bounded (Theorem on homeomorphism of Banach spaces, see [5], p. 176).

3 The random walk process and the invertibility of the difference operator

The random walk process is defined in [1], p. 47. This process is of the form (15) with the *difference operator* $\Phi(B) = \Delta = I - B$. In 2.1 we said this process could not be expressed in the form of a stationary process (1). The source of its nonstationarity is the stochastic trend-the cummulation of the elements of a white noise $\mathcal{E} = (\varepsilon_{t-n})$. The equality

$$(I - B)^{-1} = \sum_{n=0}^{\infty} B^n \tag{18}$$

has only formal meaning (see Ex. 1 in 2.1). In (18) only left side has a sense if we restrict ourselves to the spaces $\ell_p = \ell_p(L_2)$ for $1 \leq p < \infty$.

3.1 Hilbert space and adjoint operator

The dual space of a Hilbert space H , i.e. the space of all continuous linear functionals on H , is due to Riesz representation theorem (see [5], p. 233) isometrically isomorphic to H. The *adjoint operator to the operator* $A \in [H]$ is the unique operator $A^* \in [H]$ satisfying $(Au, w) = (u, A^* w)$ for any $u, w \in H$. It holds that $\|A^*\| = \|A\|$, the resolvent set and the spectrum of A and A^* are the same, however their structures can be different. If $A = A^*$ the operator A is said to be *self-adjoint.* The notion of adjoint operator can be introduced even if the operator A is unbounded with domain $D(A)$ dense in H (see [4], p. 300). The adjoint operator is always the *closed operator*, i.e. it holds

$$w_n \to w,\ A^* w_n \to w^* \Rightarrow w \in D(A^*),\ A^* w = w^* \tag{19}$$

for any sequence $(w_n) \subset D(A^*)$.

An operator A with $D(A)$ dense in H is said to be the *symmetric operator* if $(Au, w) = (u, Aw)$ for any $u, w \in D(A)$. If $D(A) = H$ it holds $A = A^*$. It means that A is closed and therefore $A \in [H]$ due to the Closed graph theorem (see [5], p. 177). The relationship between symmetric and self-adjoint operator is that the symmetric operator need not be self-adjoint but there may sometimes exist an extension to the self-adjoint operator and then this is the maximal symmetric extension of the given symmetric operator. The inverse of self-adjoint (symmetric) operator is always self-adjoint (symmetric). If a bounded operator A is not symmetric there is a simple mean how to make it to be symmetric. It suffices to form $A^* A$. This construction is the base of the well-known least square method. The operator $A^* A$ is always *nonnegative operator* (analogous to the positive semi-definite matrix) since $(A^* Au, u) = (Au, Au) \geq 0$ for any $u \in D(A)$.

3.2 The symmetrization of the random walk process

In 1.2 the Hilbert space $\ell_2 = \ell_2(L_2)$ was introduced. Its elements are, inter alia, the stationary processes up to the level N. The adjoint operator to the operator (16) is $B^* \in [\ell_2(L_2)]$. It is represented by the matrix

$$\mathbf{B}^* = \begin{bmatrix} 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{20}$$

The operators (20) and (16) have the same spectrum (see 3.1). However the structure of both spectra is different: in the case of (20) the entire of the unit circle forms the residual part while its boundary belongs to the continuous part of its spectrum. The adjoint to the difference operator $\Delta = I - B$ is the operator $\Delta^* = I - B^*$. The operator $T = \Delta^* \Delta$ is symmetric and nonnegative. It has the matrix representation

$$\mathbf{T} = \begin{bmatrix} 1 & -1 & 0 & \dots \\ -1 & 2 & -1 & \dots \\ 0 & -1 & 2 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{21}$$

The operator T is self-adjoint because it is bounded. Its inverse T^{-1} is unbounded and self-adjoint (see 3.1) and hence closed. The domain of T^{-1} is dense in $\ell_2(L_2)$, not equal to $\ell_2(L_2)$. The solution of the equation

$$T\,\mathbf{X} = (I - B^*)\mathcal{E} \tag{22}$$

is unique if the right side is in the domain of T^{-1}. Suppose $\mathcal{E} = (\varepsilon_{t-n}) \in \ell_2(L_2)$ is a white noise up to the level N. Let $\mathbf{W}_\mathrm{n} = (\boldsymbol{\varepsilon_t}, \boldsymbol{\varepsilon_{t-1}}, \cdots, \boldsymbol{\varepsilon_{t-n}}, 0,0, \cdots)$. Then $(I - B^*)\mathbf{W}_\mathrm{n} \in D(T^{-1})$ (see 2.2). The unique solution of the equation

$$T\,\mathbf{X} = (I - B^*)\mathbf{W}_\mathrm{n} \tag{23}$$

is $\mathbf{X}_n = (X_t, X_{t-1}, \cdots, X_{t-n}, X_{t-n-1}, 0,0, \cdots)$. The sequence $(\mathbf{W}_\mathrm{n}) \subset D(T^{-1})$ and $\|\mathbf{W}_\mathrm{n} - \mathcal{E}\|_2 \to 0$. If the sequence of the solutions of (23) is Cauchy, i.e. $\|\mathbf{X}_n - \mathbf{X}_m\|_2 \to 0$ for $n, m \to \infty$ then it is convergent with limit $\mathbf{X}^* \in \ell_2(L_2)$. Now we use that T^{-1} is closed and in (19) set $A^* = T^{-1}$ and $w_n = \mathbf{W}_\mathrm{n}$. Hence $\mathbf{X}^*$ is the unique solution of (22).

Conclusion

In spite of the fact the random walk process is not stationary, the equation $(I - B)\,\mathbf{X} = \mathcal{E}$ has a unique solution if we admit the white noise $\mathcal{E}$ with variance converging to zero at least from some time point at a distant past. The convergence of the solutions of finite dimensional restrictions of the original problem to its solution can be explained much easier if we employ the symmetrization of the problem. The convergence is achieved by the closedness of the inverse operator to the self-adjoint nonnegative operator $T = \Delta^*\Delta$.

References

[1] Arlt, J.: *Moderní metody modelování ekonomických časových řad*. Grada Publishing 1999

[2] Dhrymes, J.P.:*Distributed lags. Problems of Estimation and Formulation*, NORTH HOLLAND, Amsterdam 1985

[3] Horský, R.: The lag operator and its spectral properties. *Mathematical Methods in Economics 2013* Proceedings Part I (2013), 285–290.

[4] Riesz, F., Nagy, B.: *Functional analysis*. Copyright 1955 by Frederick Ungar Publishing, Co, transl. from Lecons d´analyse fonctionelle by Leo F. Boron, Dep. Of Math., University of Michigan

[5] Taylor, A.E.: *Úvod do funkcionální analýzy*. Academia, Praha 1973

A note on the use of copulas in chance-constrained programming

Michal Houda[1]

Abstract. In this paper we are concentrated on a problem of linear chance-constrained programming where the constraint matrix is considered random with a known distribution of the matrix rows. The rows are not considered to be independent; instead, we make use of the copula notion to describe the dependence of the matrix rows. In particular, the distribution of the rows is driven by so-called Archimedean class of copulas. We provide a review of very basic properties of Archimedean copulas and describe how they can be used to transform the stochastic programming problem into a deterministic problem of second-order cone programming. Also the question of convexity of the problem is explored and importance of the selected class of copulas is commented. At the end of the paper, we provide a simple example to illustrate the concept used.

Keywords: Chance-constrained optimization, Archimedean copulas, Convexity, Second-order cone programming.

JEL classification: C44
AMS classification: 90C15

1 Introduction

In many economic decision problems we are faced to the problem how to deal with uncertainty of the data. In this contribution we introduce an extension to the probabilistic programming approach based on the description of probabilistic properties of the data through the notion of copulas.

1.1 Problem Formulation

Consider an *uncertain linear optimization problem* in the form

$$\min c^T x \qquad \text{subject to} \qquad \Xi x \leq h,\ x \in X \tag{1}$$

where $\Xi \subset \mathbb{R}^n$ is a deterministic closed convex set, $c \in \mathbb{R}^n$, $h = (h_1, \ldots, h_K) \in \mathbb{R}^K$ are deterministic vectors, and $\Xi \in \mathbb{R}^K \times \mathbb{R}^n$ is an uncertain (unknown) matrix. We denote $\Xi_1^T, \ldots, \Xi_K^T$ the rows of the matrix Ξ. It is easy to see that if X is polyhedral, and a realization of the data element Ξ is known in advance, (1) is an instance of the classical linear optimization problem. As already G. Dantzig noticed in his early work [4], this is rarely the case. Instead, we have to consider *uncertainty of the data* as a natural element of the optimization model.

Several approaches can be drawn to deal with uncertainty of the data. Among well known ones we cite (ex-post) sensitivity analysis of the linear programming, parametric programming (in which data element is considered to be parameter), or robust approach (the constraints of the problem are required to be satisfied for all realizations from some predefined uncertainty set). In our paper, we concentrate on the stochastic programming approach: we suppose that Ξ is a random vector with a known distribution, and the constraints are required to be satisfied with a prescribed, sufficiently high probability. More specifically, for a fixed $p \in [0; 1]$, we consider a *chance-constrained linear optimization problem* of the form

$$\min c^T x \qquad \text{subject to} \qquad \Pr\{\Xi x \leq h\} \geq p,\ x \in X. \tag{2}$$

[1]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, houda@ef.jcu.cz

For easier referencing, we drop the deterministic constraint $x \in X$ from our future consideration (without losing the generality of our conclusions), and denote the feasible set of problem (2) as

$$X(p) := \big\{x \in \mathbb{R}^n \mid \Pr\{\Xi x \leq h\} \geq p\big\}. \tag{3}$$

Chance-constrained programming problems were treated many times in the literature, starting with the work [2], and summarized by the classical book [14] and recent chapters [15] and [5]. It were realized that problem (2), especially in the presented form with random data matrix, is not easy to analyze — from the theoretical as well as from the practical point of view. It is worth to declare that two main difficulties were revealed: first, even with its inner linear structure, the feasible set $X(p)$ has not to be convex; and second, the presence of multidimensional integrals prevents the probability in (2) to be easily calculated. We concentrate on the first issue here. Convexity of the set $X(p)$ was treated several times in the literature; apart from mentioned references we mention recent results by [6], [7], and [16]. In our paper, we will use an approach developed by [11] and provide conditions under which the problem happen to be convex, together with a simple application illustrating the concept used.

2 Dependence — Notion of Copulas

Convexity results mentioned shortly in the introduction of the paper (except the last one) are based on the assumption of *row independence*, that is, the random vectors $\Xi_1^T, \ldots, \Xi_K^T$ are supposed to be pairwise independent. Although it can be seen as an obstacle, we have a handy tool to generalize this assumption: a copula notion. This is well known from the field of probability theory and mathematical statistics but rarely used in stochastic optimization, especially it was not used to describe structural properties of chance-constrained problems. Apart of already mentioned reference we can cite [8] and [1] but these works do not directly concern the problem of the form (2). In this section we provide necessary definitions and theorems needed to describe dependence properties of problem (2); we refer the reader to the book [13] for details.

Definition 1. A *copula* is the distribution function $C : [0;1]^K \to [0;1]$ of some K-dimensional random vector whose marginals are uniformly distributed on [0;1].

It is Sklar's Theorem (given below) that justifies usefulness of copula notion to describe dependence properties of a problem. Before stating it, we recall the definition of quantile function which plays a fundamental role in the sequel of the paper.

Definition 2. Given a one-dimensional distribution function $F : \mathbb{R} \to [0;1]$, its *quantile function* is defined as

$$F^{(-1)}(\pi) := \inf\{t \in \mathbb{R} \mid F(t) \geq \pi\}.$$

Proposition 1 (Sklar's Theorem)**.** *For any K-dimensional distribution function $F : \mathbb{R}^K \to [0;1]$ with marginals $F_1, \ldots, F_K$, there exists a copula C such that*

$$\forall z \in \mathbb{R}^K \quad F(z) = C(F_1(z_1), \ldots, F_K(z_K)). \tag{4}$$

If, moreover, F_k are continuous, then C is uniquely given by

$$C(u) = F(F_1^{-1}(u_1), \ldots, F_K^{-1}(u_K)). \tag{5}$$

Otherwise, C is uniquely determined on $\operatorname{range} F_1 \times \cdots \times \operatorname{range} F_K$.

Through Sklar's Theorem, an arbitrary dependence structure can be described by an appropriately chosen (unique) copula. A leading example is so-called *independent (product) copula*

$$C_\Pi(u) = \prod_k u_k \tag{6}$$

providing us the well known independence formula for distributions functions — if the marginals $F_1, \ldots, F_K$ are independent then their joint distribution function F is given by

$$F(z) = C_\Pi\big(F_1(z_1), \ldots, F_K(z_K)\big) = \prod_{k=1}^{K} F_k(z_k).$$

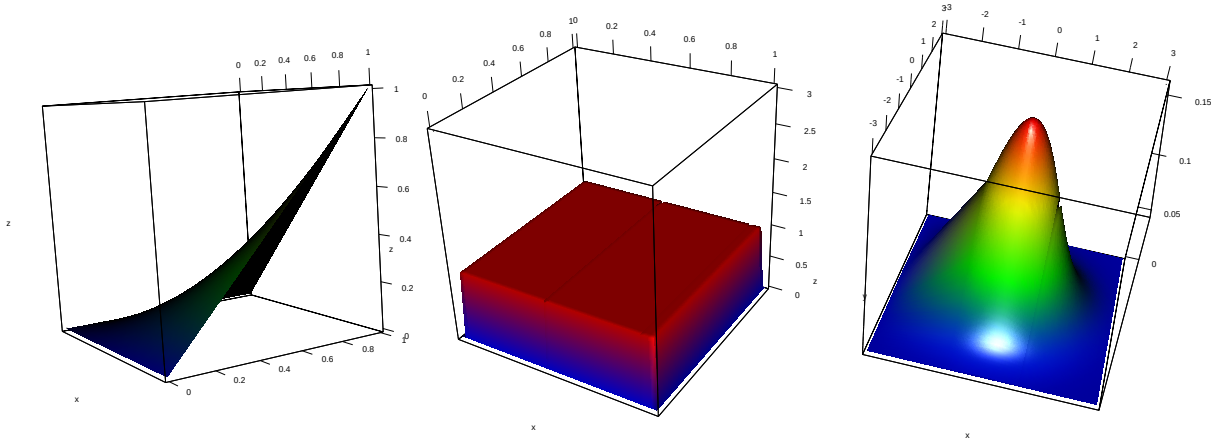


Figure 1 Independent copula: distribution, density, and density with standard normal marginals.

For reader convenience we provide a graphical representation of the independence copula in Figure 1. The third graph of the figure is the well-known two-dimensional density of the normal distribution with independent marginals. The contours of this density are in form of "regular" ellipsoids (circles in this special case).

To an easy treatment of problem (2) we need an algebraically convenient class of copulas. This is provided by the Archimedean copulas.

Definition 3. A copula C is called *Archimedean* if there exists a continuous strictly decreasing function $\psi : [0; 1] \to [0; +\infty]$, called *generator of* C, such that $\psi(1) = 0$ and

$$C(u) = \psi^{(-1)} \left(\sum_{i=1}^{n} \psi(u_i) \right). \tag{7}$$

If $\lim_{t\to 0} \psi(t) = +\infty$ then C is called a *strict Archimedean copula* and ψ is called a *strict generator*.

A necessary and sufficient condition for a continuous, strictly decreasing function to be an Archimedean generator is not easy task and was provided by [12]. We need only a sufficient condition for strict Archimedean generators.

Definition 4. A real function $f : \mathbb{R} \to \mathbb{R}$ is called *completely monotonic* on an open interval $I \subseteq \mathbb{R}$ if it is nondecreasing, differentiable for each order k, and its derivatives alternate in sign, i. e.

$$(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k} f(t) \geq 0 \ \forall k = 0, 1, \ldots, \text{ and } \forall t \in I. \tag{8}$$

Proposition 2. *Let $\psi : [0; 1] \to \mathbb{R}_+$ be a strictly decreasing function with $\psi(1) = 0$ and $\lim_{t\to 0} \psi(t) = +\infty$. Then it is a generator of a strict Archimedean copula for each dimension $K \geq 2$ if and only if $\psi^{(-1)}$ is completely monotonic on $(0; +\infty)$.*

It is easy to see that the independent copula is Archimedean with the strict generator $\psi(t) = -\ln t$. Another popular example of an Archimedean copula is the *Gumbel–Hougaard copula* defined by $\psi(t) = (-\ln t)^\theta$ where $\theta \geq 1$ can be seen as a parameter of the dependence. The Gumbel–Hougaard copula is presented in Figure 2. Compared to the independent case, the density (the third subfigure) is "pointed" and the ellipsoidal contours of the independent copula take form of drops. For other illustrations of Archimedean copulas (Clayton, Joe) see [11]; an exhaustive listing of 22 families of Archimedean copulas treated by the literature is given by Table 4.1 in [13].

3 Convex Reformulation

The Archimedean copulas can be used for an equivalent convex reformulation of the feasible set $X(p)$.

Theorem 3 ([11], Theorem 3.1)**.** *Suppose, in* (3)*, that the rows Ξ_k^T have n-variate normal distribution with means μ_k and positive definite covariance matrices Σ_k. Then the feasible set of problem* (2) *can be*

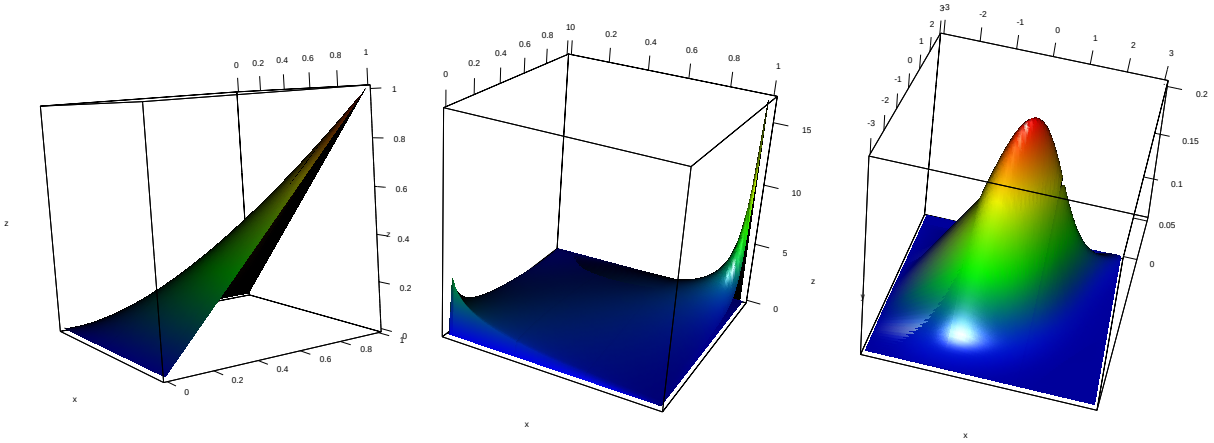


Figure 2 Gumbel-Hougaard copula ($\theta = 1.6$): distribution, density, and density with standard normal marginals.

equivalently written as

$$X(p) = \left\{ x \mid \exists y_k \geq 0 : \sum_k y_k = 1,\ {\mu_k}^T x + \Phi^{(-1)}\left(\psi^{(-1)}(y_k \psi(p))\right) \sqrt{x^T \Sigma_k x} \leq h_k\ \forall k \right\} \tag{9}$$

where Φ is the distribution function of a standard normal distribution and ψ is the generator of an Archimedean copula describing the dependence properties of the rows of the matrix Ξ. Moreover, if $\psi^{(-1)}$ is completely monotonic, and $p > p^* := \Phi\left(\max\{\sqrt{3}, 4\lambda_{max}^{(k)}[\lambda_{min}^{(k)}]^{-3/2}\|\mu_k\|\}\right)$, where $\lambda_{max}^{(k)}, \lambda_{min}^{(k)}$ are the largest and lowest eigenvalues of the matrices Σ_k, then the problem is convex.

Proof. See [10]. The threshold p^* was provided for independent case by [7] using properties of r-concave functions (see also [9]) but it have not to be changed under our dependency assumptions. □

Theorem 3 offers a way to replace multidimensional integral in (3) by a set of one-dimensional (separable) constraints of the form (9). Reference [11] (based on the work [3]) also provides a numerical approximation of (9) by two second-order cone programming (SOCP) problems, using the fact that the function $y \mapsto H(y) := \Phi^{(-1)}\left(\psi^{(-1)}(y_k \psi(p))\right)$ is convex. For example, a piecewise-convex approximation of this functions leads to the following SOCP problem giving a lower bound for the optimal value of the original probabilistic problem

$$\begin{aligned} \min c^T x \qquad & \text{subject to} \\ & {\mu_k}^T x + \sqrt{z^{kT} \Sigma_k z^k} \leq h_k,\ z^k \geq a_{kj} x + b_{kj} w^k\ (\forall k, \forall j) \\ & \sum_k w^k = x,\ w^k \geq 0,\ z^k \geq 0\ (\forall k), \end{aligned} \tag{10}$$

where

$$a_{kj} := H(y_{kj}) - b_{kj} y_{kj}, \qquad b_{kj} := \frac{\psi(p)}{\phi(H(y_{kj}))\psi'\left(\psi^{(-1)}(y_{kj}\psi(p))\right)},$$

y_{kj} are given partition points of the interval $(0; 1]$, and ϕ is the standard normal density. In [11], also an inner approximation is provided for the problem.

4 Application to the Model with EEA Indicators

To present a simple example of using this approach, we pursue the direction presented already in [9] and previous works referenced therein. We briefly summarize the model and its description; the details can be found in the mentioned reference [9]. We did not include numerical results (based on simulations) in the current paper as they are matter of our current and still not finished research.

Consider vector mappings $i_k(x;\xi) := x^T A^k \xi$, representing a functional dependence of the EEA indicators (of ecological stability) i_k on decision vector x (possible ecological arrangements), and uncertainty factor ξ. The indicators of ecological stability monitor basic properties of the environment, e.g. air and water quality, pollution, noise, etc. Possible arrangements include desulphurisation of plants, water cleaners, soundproofing of roads, limits provided by law, and many others.

To convene our assumption we suppose this dependence to be linear. Denote L_k required threshold values for the indicators i_k; apart x, L_j are also considered to be decision variables of the problem in our formulation. Our aim is to maximize the ecological limits in order to meet them at least with the required fixed probability p (i. e., to find the optimal critical values for ecological indicators achievable with high probability). The problems can be then formulated as

$$\max \sum_k L_k \qquad \text{subject to} \qquad \Pr\{x^T A^k \xi \geq L_k, k \in \mathcal{K}\} \geq p,\ x \in X_0; \tag{11}$$

X_0 contains budget and other deterministic constraints.

Problem (11) falls into the frame of problem (2). Suppose that the joint distribution of the constraints rows $\mathcal{J}$ is driven by an Archimedean (for example, by the Gumbel–Hougaard) copula. If ξ is normally distributed random vector, then so is the product $A^k\xi$, and the problem is convex by Theorem 3 for sufficiently high value of p. The outer approximating SOCP problem of the form (10) is straightforward, and lower bound for the optimal value can be easily found as the side effect of the problem. The parameters μ_k, Σ_k are provided by probabilistic properties of $A^k\xi$. Using the exact form of the generator ψ of the Gumbel–Hougaard copula, we can provide the coefficients a_{kj}, b_{kj} in the following explicit form

$$a_{kj} := \Phi^{(-1)}\left(p^{y_{kj}^{1/\theta}}\right) - b_{kj} y_{kj}, \qquad b_{kj} := \frac{(-\ln p)^\theta p^{y_{kj}^{1/\theta}}}{\theta\phi\left(\Phi^{(-1)}\left(p^{y_{kj}^{1/\theta}}\right)\right)\left(\ln p^{y_{kj}^{1/\theta}}\right)^{\theta-1}}.$$

5 Conclusion

In this work we have provided a methodological note on a stochastic programming problem with chance constraints. In particular, a use of a specific class of Archimedean copulas having completely monotonic inverse generators on a problem with normally distributed rows was given. The problem is reformulated appropriately and the convexity of the feasible set is stated for a sufficiently high probability level. We also describe how the problem can be approximated by an second-order cone programming problem which is solvable by traditional optimization procedures. The accompanying (but still artificial) example on indicators of ecological stability offers an insight into a possible application of the methodology described.

Acknowledgements

This work was partially supported by the grant No. 13-14445S of the Czech Science Foundation.

References

[1] van Ackooij, W.: Eventual convexity of chance constrained feasible sets. *Optimization*, 2013. Published online.

[2] Charnes, A., Cooper, W. W., and Symonds, G. H.: Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management Science* **4**(3) (1958), 235–263.

[3] Cheng, J. and Lisser, A.: A second-order cone programming approach for linear programs with joint probabilistic constraints. *Operations Research Letters* **40**(5) (2012), 325–328.

[4] Dantzig, G. B.: Linear programming under uncertainty. *Management Science* **1** (1955), 197–206.

[5] Dentcheva, D.: Optimization models with probabilistic constraints. In *Lectures on Stochastic Programming: Modeling and Theory* (A. Shapiro, D. Dentcheva, and A. Ruszczyński, eds), volume 9 of *MOS-SIAM Series on Optimization*, SIAM, Philadelphia, 2009, 87–153.

[6] Henrion, R.: Structural properties of linear probabilistic constraints. *Optimization*, **56**(4) (2007), 425–440.

[7] Henrion, R., and Strugarek, C.: Convexity of chance constraints with independent random variables. *Computational Optimization and Applications* **41**(2) (2008), 263–276.

[8] Henrion, R., and Strugarek, C.: Convexity of chance constraints with dependent random variables: the use of copulae. In *Stochastic Optimization Methods in Finance and Energy* (M. Bertocchi, G. Consigli, and M. A. H. Dempster, eds), volume 163 of *International Series in Operations Research & Management Science*, Springer, New York, 2011, 427–439.

[9] Houda, M.: Convexity in stochastic programming model with indicators of ecological stability. In *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012* (Ramík, J., and Stavárek, D., eds.), Silesian University, School of Business Administration, Karviná, 2012, 314–319.

[10] Houda, M., and Lisser, A.: Second-order cone programming approach for elliptically distributed joint probabilistic constraints with dependent rows. Research Report 1566, Laboratoire de Recherche en Informatique, Université Paris Sud XI, October 2013.

[11] Houda, M., and Lisser, A.: On the use of copulas in joint chance-constrained programming. In *Proceedings of the 3rd International Conference on Operations Research and Enterprise Systems*, 2014, 72–79.

[12] McNeil, A. J., and Nešlehová, J.: Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *Annals of Statistics* **37**(5B) (2009), 3059–3097.

[13] Nelsen, R. B.: *An Introduction to Copulas.* Springer, New York, 2nd edition, 2006.

[14] Prékopa, A.: *Stochastic Programming.* Akadémiai Kiadó, Budapest, 1995.

[15] Prékopa, A.: Probabilistic programming. In *Stochastic Programming* (A. Ruszczyński and A. Shapiro, eds), volume 10 of *Handbooks in Operations Research and Management Science*, Elsevier, Amsterdam, 2003, 267–352.

[16] Prékopa, A., Yoda, K., and Subasi, M. M.: Uniform quasi-concavity in probabilistic constrained stochastic programming. *Operations Research Letters* **39**(3) (2011), 188–192.

Multiple Criteria Approaches in Insurance Fraud and Receivables Nonpayment Risk Modeling

Eduard Hozlár[1], Vladimír Mlynarovič[2]

Abstract. The paper combines outranking and optimization approaches to insurance fraud and receivables nonpayment risk modeling. Methodology of criteria scoring and weighting is based on a definition the fraud (risk) intensity function for each assumed criterion that described intensity level for possibility of the fraud (nonpayment). The values of this function weighted on the base of rules importance will provide outflow for each claim. The outflow is confronted with investigation level. In ideal case one will have such selections of intensity function, weight of rules importance and investigation level, where each realized investigation will identify the fraud (nonpayment0 and there will be no fraud (nonpayment) among claims that were not investigated.

Keywords: insurance fraud, receivables, rule fraud or risk intensity function, investigation level, claim outflow, linear programming problem, goal programming.

JEL Classification: C61
AMS Classification: 90C29

1 Introduction

The paper combines outranking ([1], [3]) and optimization approaches ([4], [5]) to insurance fraud and receivables nonpayment risk modeling. The aim of the paper is to develop a normalized model approach for identification of potential insurance frauds and potential nonpayment of receivables. The suggested methodology starts with the definition of a rule intensity function. Methodology of rules (criteria) scoring and weighting is based on this definition of the fraud (risk) intensity function for each assumed criterion that described intensity level for possibility of the fraud (nonpayment). The values of this function weighted on the base of rules importance will provide outflow for each claim. Optimization models provide tools for a classification of claims from a viewpoint of a fraud (nonpayment). In the models the outflow is confronted with investigation level. In ideal case one will have such selections of intensity function, weight of rules importance and investigation level, where each realized investigation will identify the fraud (nonpayment) and there will be no fraud (nonpayment) among claims that were not investigated. In the final part of the paper some typical rule intensity function are introduced.

2 The methodology

Let R is the set of *rules* (criteria) and F is the set of rule fraud (risk) intensity functions, where for each $r \in R$ we have selected function $f_r \in F$ which for selected rule value d specifies

$$f_r(d) \in \langle 0,1 \rangle$$

where

$$f_r(d)\begin{cases} = 0, & \textit{zero possibility of fraud (nonpayment)} \\ \approx 0, & \textit{low possibility of fraud (nonpayment)} \\ \approx 1, & \textit{high possibility of fraud (nonpayment)} \\ = 1, & \textit{strict possibility of fraud (nonpayment)} \end{cases}$$

If w_r is the importance weight for rule r, where

$$\sum_{r \in R} w_r = 1, \quad w_r \geq 0, \forall r$$

then for a claim i we have its *outflow*

[1] Comenius University, FSEV, Institute of Economics, Slovakia, eduard.hozlar@fses.uniba.sk.
[2] Comenius University, FSEV, Institute of Economics, Slovakia, vladimir.mlynarovic@fses.uniba.sk.

$$O_i = \sum_{r \in R} w_r \, f_r(d_i) \in \langle 0,1 \rangle$$

where d_i is value rule for claim *i.*

2.1 Settings

Suppose we have from a history the set I of claims that consists of the two subsets I_1 and I_2, where

$$I = I_1 \cup I_2, \quad I_1 \cap I_2 = \varnothing$$

and subset I_1 contains all claims without a fraud (nonpayment) and the subset I_2 contains all claims where a fraud (nonpayment) was confirmed. Let

$$l \in \langle 0,1 \rangle$$

is so called *investigation level* in the sense that all claims with outflow under the level will not be investigated and all claims with outflow at least on the investigation level should be investigated. It means that we need such settings (rule function selections, weight of importance selections and investigation level selections) for which

$$O_i < l, i \in I_1$$
$$O_i \geq l, i \in I_2$$

Linear programming approach

Suppose that each of the rules is managed by an expert of the rule which is able to select the rule fraud (risk) intensity function from an existing list (a possible starting list is presented in the part 2.2), or specified a new function for the rule. If there is no imagination on investigation level l, the following linear programming problem can be used to find corresponding weights of rules importance and investigation level:

$$\min \quad l$$

subject to

$$O_i \leq l - \varepsilon, i \in I_1$$
$$O_i \geq l, i \in I_2$$
$$\sum_r w_r = 1, \; w_r \geq 0, \; \forall r$$

where ε is a sufficiently small number.

Goal programming approach

If there is an imagination on investigation level l, e.g. from a historical analysis, then the following goal programming problem can be used to find importance weights for rules:

$$\min \quad \alpha$$

subject to

$$O_i - \alpha \leq l - \varepsilon, i \in I_1$$
$$O_i + \alpha \geq l, i \in I_2$$
$$\sum_r w_r = 1, \; w_r \geq 0, \; \forall r, \alpha \geq 0$$

where α is so called minmax variable. In the both cases the duality theory will be used to find a critical value of parameters

A rule (criterion) importance

An importance of each rule j can be identified on the base of following linear programming problem solution

$$\min \quad w_j$$

subject to

$$O_i \le l - \varepsilon, i \in I_1$$
$$O_i \ge l, i \in I_2$$
$$\sum_r w_r = 1,\ w_r \ge 0,\ \forall r, \alpha \ge 0$$

where zero optimal value of the weight says that the rule has no effect on the fraud (nonpayment) identification.

2.2 Rule intensity functions

As a starting list of such functions we suggest the following six rule fraud (risk) intensity functions:

1.
$$f(d) = \begin{cases} 0, if\ d = 0 \\ 1, if\ d > 0 \end{cases}$$

2.
$$f(d) = \begin{cases} 0, if\ d \le q \\ 1, if\ d > q \end{cases}$$

3.
$$f(d) = \begin{cases} d / p, if\ d \le p \\ 1, if\ d > p \end{cases}$$

4.
$$f(d) = \begin{cases} 0 & , if\ d \le q \\ 0.5, if\ d > q, d \le p \\ 1 & , if\ d > p \end{cases}$$

5.
$$f(d) = \begin{cases} 0 & , if\ d \le q \\ \dfrac{d-q}{p-q}, if\ d > q, d \le p \\ 1 & , if\ d > p \end{cases}$$

6.
$$f(d) = 1 - \mathrm{Exp}(-(\mathrm{d}^2) / (2\,\mathrm{s}^2))$$

where q is zero intensity threshold, p is strong intensity threshold and s is the standard deviation for *Gaussian* function. In life insurance fraud modeling, for example, such rules can be formulated from viewpoint of agent, doctor (hospital), diagnosis, region, etc., and could be derived from number of new claims registered per month, sum of payments of claims paid per month, sum of declined payment of claims paid per month, number of claims of insured person less then 10 (15) years old, number of days between policy beginning and date of the claim, number of hospitalization days on claims, number of payments 5% under authorization limit, number of claims in which date of occurrence or outpatient date is Saturday or Sunday or it happen nightly, number of outpatient treatment (diagnosis/hospitalization) at the same date, and so on.

For a production company that has many customers, the intensity of the risk of non-payment for delivered goods may be determined by the following parameters ([2]):
Indicators proportionally increasing the risk of non-payment:

- the size of the debt to the state (social insurance, taxes, health insurance);
- the size of the debt own to financial institutions (banks, leasing, insurance);
- the size in trade payables.

Indicators the growth of which reduces the risk of non-payment:

- sales;
- profitability;
- equity;
- growth prospects.

The practical application of this, in authors' opinion, new methodology requires a robust analytical system, e.g. SAS, including its optimization modules. For example, in insurance fraud, an identification of investigation level could require to examine claims counted in millions. As a result this approach provides a normalized picture of all flows in the modeled system with a possibility to compare various aspects of the claims and identify key sources of frauds and nonpayment. This normalization in the suggested approach is its main advantage in comparison with approaches that are authors known from practice in some insurance companies. As a by product we also have a rating of agents, doctors (hospitals), diagnosis, …, in life insurance modeling and clients of factoring company in nonpayment modeling.

References

[1] Brans, J. P., Mareschal, B. and Vincke, Ph.: A new family of outranking methods in multicriteria analysis. In: Brans, J.P. (ed): Operational Research '84, North Holland 1984, 477 – 490.
[2] Daniel, P.: *Správa a vymáhanie pohľadávok*. IURA Edition, Bratislava, 2013.
[3] Mlynarovič, V.: *Modely a metódy viackriteriálneho rozhodovania*. Ekonóm., Bratislava, 1998.
[4] Mlynarovič, V. a Hozlár, E.: *Viackrtiteriálne rozhodovanie*. Ekonomická univerzita, Bratislava, 1993.
[5] Steuer, R. E.:Multiple criteria optimization: Theory, Computation, and Application0John Wiley & Sons, Inc. New York, 1986.

Regression among parts of compositional data with an economic application

Klára Hrůzová[1], Karel Hron[2], Peter Filzmoser[3], Valentin Todorov[4]

Abstract. Compositional data follow the Aitchison geometry on the simplex. Therefore, it is not appropriate to apply regression analysis directly to the raw compositional data. For this purpose, isometric logratio (ilr) coordinates were formed to express compositional data in coordinates with respect to an orthonormal basis on the simplex. For regression analysis between the response part and an explanatory subcomposition such coordinates are required that enable assigning coordinates to both the response variable and one of the covariates simultaneously. Moreover, as all coordinates contain measurement errors, the use of ordinary least squares regression for parameter estimation would lead to biased results. Therefore, orthogonal regression is used leading to an intuitive geometric interpretation. In the application part, the relation between different activities of gross value added (that form inherently compositional data) are investigated in ilr coordinates.

Keywords: isometric logratio transformations, orthogonal regression, principal component analysis.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Compositional data analysis deals with multivariate observations carrying only relative information [1]. They are often represented in proportions or percentages, and thus have a certain constant sum constraint, like 1 or 100. Because of the different sample space of compositions, the simplex $\mathcal{S}^D$, the Aitchison geometry with Euclidean vector space properties was introduced [2]. Nevertheless, as the D-part simplex has dimension $D-1$, there exists no canonical basis in the Aitchison geometry. Due to specific features of compositions, captured by the Aitchison geometry, it is not appropriate to directly apply standard statistical methods to compositional data. The only way how to analyze compositions reasonably is to express them in orthonormal coordinates with respect to the Aitchison geometry before applying the standard statistical procedures. In some cases it is also possible to go back to the original space of compositional data to allow for an interpretation in the original space.

Isometric logratio (ilr) transformations [2], that result in orthonormal coordinates with respect to the Aitchison geometry, are often used for their good theoretical properties and possibly also an easy interpretability. As there exist infinitely many orthonormal bases in the Aitchison geometry, we can choose from many possible options how to form the ilr (orthonormal) coordinates. The concrete choice always depends on good interpretability of outputs of statistical processing. Apart from special cases, where an a priory information can be used to form interpretable orthonormal coordinates, we seek for an automated version of coordinates that could be used for most problems that occur in practice.

[1]Department of Mathematical Analysis and Applications of Mathematics, Department of Geoinformatics - Faculty of Science, Palacký University, 17. listopadu 12, Olomouc, Czech Republic, klara.hruzova@gmail.com
[2]Department of Mathematical Analysis and Applications of Mathematics, Department of Geoinformatics - Faculty of Science, Palacký University, 17. listopadu 12, Olomouc, Czech Republic, hronk@seznam.cz
[3]Department of Statistics and Probability Theory, Vienna University of Technology, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria; Department of Geoinformatics - Faculty of Science, Palacký University, 17. listopadu 12, Olomouc, Czech Republic, p.filzmoser@tuwien.ac.at
[4]United Nations Industrial Development Organization (UNIDO), Vienna International Centre, Wagramerstr. 5, Vienna, Austria, valentin.to@gmail.com

Suppose we have a composition of D parts, $\mathbf{x} = (x_1, \dots, x_D)'$, then we define ilr coordinates as

$$z_i^{(l)} = \sqrt{\frac{D-i}{D-i+1}} \ln \frac{x_i^{(l)}}{\sqrt[D-i]{\prod_{j=i+1}^{D} x_j^{(l)}}},\ i = 1, \dots, D-1, \tag{1}$$

where $\mathbf{z}_l = (z_1^{(l)}, \dots, z_{D-1}^{(l)})',\ l = 1, \dots, D$, is a real vector and $(x_1^{(l)}, \dots, x_1^{(l)}, x_{l+1}^{(l)}, \dots, x_D^{(l)})'$ stands for a permutation of the parts of $\mathbf{x}$ where always the l-th compositional part occupies the first position, $(x_l, x_1, \dots, x_{1-1}^{(l)}, x_{l+1}^{(l)}, \dots, x_D)'$. Thus the first ilr variable $z_1^{(l)}$ explains all the relative information about the original part x_l (through the logratio of the part of interest to the remaining parts, expressed by their geometric mean) and the coordinates $z_2^{(l)}, \dots, z_{D-1}^{(l)}$ explain the remaining logratios in the composition. Note that from this construction, the coordinate $z_1^{(l)}$ cannot be assigned directly with the corresponding part x_l, it represents only a measure of dominance of the part x_l with respect to the other parts.

When we are interested in the analysis of the relations between different parts of a composition, one natural problem arises as at least two compositional parts (the response part x_l and explanatory part(s)) are of simultaneous interest, although their positions in the regression model are different. The solution is then to express the response variable using $z_1^{(l)}$ from (1), which contains information about the relative relations (ratios) of x_l to all remaining parts of the composition [4]. Consequently, we can proceed to a coordinate representation of the explanatory subcomposition $(x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_D)'$. For this purpose, we use coordinates $z_2^{(k)}, \dots, z_D^{(k)}$ corresponding to the reordered subcomposition $(x_k, x_1, \dots, x_i, \dots, x_D)'$, $i \neq \{k, l\},\ k = 1, \dots, D,\ k \neq l$. Then the coordinate $z_2^{(k)}$ explains all the relative information about the part x_k in the resulting subcomposition. Considering the range of k, we obtain $D-1$ regression models assigned to single explanatory compositional parts. In these models only the absolute term parameter and the parameter corresponding to coordinate $z_2^{(k)}$ are interesting for further interpretation purposes of the regression analysis.

Due to the fact that both the response and the explanatory variables come from one composition, it is necessary to realize that all the variables (including the explanatory variables) are subject to errors. Therefore, the use of the ordinary least squares (OLS) regression model would lead to inconsistent estimates of the regression parameters. Thus we propose to apply the orthogonal regression model (or, equivalently, the total least squares model), a specific type of errors-in-variable (EIV) models [3], for this purpose.

2 Orthogonal regression

For simplification of the notation, we denote the matrix of n replicates of the vector $(z_2^{(k)}, \dots, z_{D-1}^{(k)})$, for a chosen $k = 1, \dots, D,\ k \neq l$, as $\mathbf{X} \in \mathbb{R}^{n \times D-2}$, and $\mathbf{Y} \in \mathbb{R}^n$ the observation vector of the response coordinate $z_1^{(l)}$. The total least-squares (TLS) method was originally introduced to solve overdetermined systems of equations $\mathbf{XB} \approx \mathbf{Y}$, where $\mathbf{X}$ and $\mathbf{Y}$ are given data (expressed in orthonormal coordinates) and $\mathbf{B} \in \mathbb{R}^{D-2}$ is the vector of unknown parameters. There is no exact solution; particularly in the case of $n > D-2$, we are seeking for an approximation.

The classical TLS problem [5] is looking for the minimal errors $\boldsymbol{\varepsilon}_X, \boldsymbol{\varepsilon}_Y$ (in the sense of the Frobenius norm, denoted by the subscript F in the following) on the given data $\mathbf{X}, \mathbf{Y}$ that make the system of equations $\widehat{\mathbf{X}}\mathbf{B} = \widehat{\mathbf{Y}},\ \widehat{\mathbf{X}} = \mathbf{X} + \boldsymbol{\varepsilon}_X,\ \widehat{\mathbf{Y}} = \mathbf{Y} + \boldsymbol{\varepsilon}_Y$ solvable, i.e.

$$\{\widehat{\mathbf{B}},\ \boldsymbol{\varepsilon}_X,\ \boldsymbol{\varepsilon}_Y\} := \operatorname{argmin}_{\boldsymbol{\varepsilon}_X, \boldsymbol{\varepsilon}_Y} \|[\boldsymbol{\varepsilon}_X,\ \boldsymbol{\varepsilon}_Y]\|_F\,, \tag{2}$$

subject to $(\mathbf{X} + \boldsymbol{\varepsilon}_X)\mathbf{B} = \mathbf{Y} + \boldsymbol{\varepsilon}_Y$. The solution is a maximum likelihood estimator $\widehat{\mathbf{B}}$ in the optimally corrected EIV model $\widehat{\mathbf{X}}\widehat{\mathbf{B}} = \widehat{\mathbf{Y}}$, $\widehat{\mathbf{X}} = \mathbf{X} + \boldsymbol{\varepsilon}_X,\ \widehat{\mathbf{Y}} = \mathbf{Y} + \boldsymbol{\varepsilon}_Y$, if the usual assumptions are fulfilled, namely that $\operatorname{vec}[\boldsymbol{\varepsilon}_X, \boldsymbol{\varepsilon}_Y]$ ("vec" forms one vector, composed of the columns of the matrix in the argument) has zero mean, and is a normally distributed random vector with a covariance matrix that is a multiple of the identity. In the EIV model, $\mathbf{X},\ \mathbf{Y}$ are the "true" data, $\widehat{\mathbf{B}}$ is the "true" value of the parameter $\mathbf{B}$ and $\boldsymbol{\varepsilon}_X,\ \boldsymbol{\varepsilon}_Y$ are the "measurement errors".

From the methodological point of view, the singular value decomposition is applied to $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}] = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}'$, where $\boldsymbol{\Sigma} = \operatorname{Diag}(\sigma_1, \dots, \sigma_{D-1})$ and $\sigma_1 \geq \dots \geq \sigma_{D-1} \geq 0$ are the singular values of $\mathbf{Z}$. Let us

define the partitions

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & v_{22} \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma_1} & \mathbf{0} \\ \mathbf{0} & \sigma_D \end{bmatrix},$$

where the matrices $\mathbf{V}_{11}$ and $\mathbf{\Sigma}_1$ are of type $(D-2) \times (D-2)$. Then a TLS solution exists iff v_{22} is non-singular; moreover, it is unique iff $\sigma_{D-2} \neq \sigma_{D-1}$. In this case it is given by

$$\widehat{\mathbf{B}} = -\mathbf{V}_{12}/v_{22} \tag{3}$$

and the corresponding TLS error matrix equals $\boldsymbol{\varepsilon}_Z = [\boldsymbol{\varepsilon}_X, \boldsymbol{\varepsilon}_Y] = -\mathbf{U}\mathrm{Diag}(\mathbf{0}, \sigma_{D-1})\mathbf{V}'$ [5]. Thus, when a unique solution $\widehat{\mathbf{B}}$ exists, it is computed from the scaled right singular vector corresponding to the smallest singular value.

It is well-known that the matrices $\mathbf{\Sigma}$ and $\mathbf{V}$ from SVD of the centered explanatory and response variables correspond to outputs of an eigenvalue decomposition of the covariance matrix, performed within principal component analysis (PCA). Thus, except for the intercept term in the orthogonal regression model, the same results as above in (3) can be obtained also using the smallest eigenvalue and the corresponding eigenvector (loading vector). We will follow this finding further in the paper.

3 Geometrical motivation

As mentioned in the previous section, the TLS (orthogonal regression) estimates of the parameters can be obtained by means of principal component analysis. We apply the general findings to a four part composition that is used in the next section, and that can be easily illustrated geometrically. For this purpose, we suppose to have a random vector $\mathbf{z} = (z_1, z_2, z_3)'$ (an orthonormal coordinate representation of the composition) and our task is to find a relationship between the response variable z_3 and covariates z_1, z_2, expressed in the form $z_3 = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \varepsilon$ for a vector of regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$.

From the geometrical point of view, the basic idea is to fit a plane to the data using PCA. The loadings of the first two principal components define a basis of the plane. As the third principal component is orthogonal to the previous ones, its loadings define the normal vector of the plane (denoted as $\mathbf{n} = (n_1, n_2, n_3)$). The plane passes through the point $\mathbf{t}$, representing the location estimate of the $n \times 3$ data matrix $\mathbf{Z}$ (the arithmetic mean in the classical case), and its perpendicular distance from the origin is $\mathbf{t}'\mathbf{n}$. The perpendicular distance from each point in $\mathbf{Z}$ to the plane (the norm of the residuals) is the inner product of each centered point and the normal of the plane. The fitted plane minimizes the sum of squared errors.

The estimated parameters $\boldsymbol{\beta}$ are obtained using the values of the normal vector, namely

$$\widehat{\beta}_0 = \frac{\mathbf{t}'\mathbf{n}}{n_3}, \ \widehat{\beta}_1 = -\frac{n_1}{n_3}, \ \widehat{\beta}_2 = -\frac{n_2}{n_3}.$$

4 Economic application

The theoretical considerations from the previous sections will be applied to an economic data set in order to investigate the relations between different activities of gross value added. According to [6] the gross value added is defined as the difference between gross output and intermediate consumption and can be decomposed into the following components: compensation of employees, gross operating surplus, mixed income and other taxes on production less subsidies on production. It can also be derived as the difference between gross domestic product and taxes on products less subsidies on products. The gross values added (GVA) can be decomposed by these activities:

- Agriculture (consisting of agriculture, forestry, hunting and fishing);
- Manufacturing[1];

[1] Manufacturing is defined as the physical or chemical transformation of materials of components into new products.

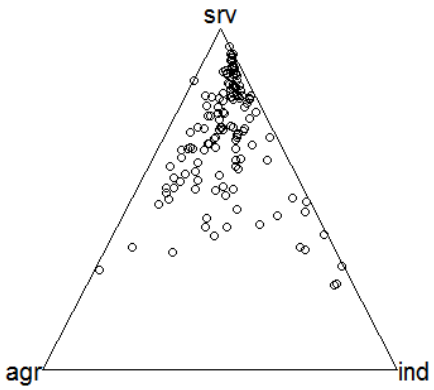


Figure 1 Ternary diagram of explanatory variables.

- Other industry (consisting of mining and quarrying; electricity, gas, steam and air conditioning supply; water supply; sewerage, waste management and remediation activities; construction);
- Services (consisting of education, health and other personal services; public administration and defense).

Thus, the GVA can be expressed as the sum of these four activities; the goal of the study is to analyze the relation between manufacturing and the rest of the activities, i.e.

$$\mathbf{Y}_{\text{manufacturing}} \sim \mathbf{X}_{\text{agriculture}} + \mathbf{X}_{\text{other industry}} + \mathbf{X}_{\text{services}}.$$

Our dataset comes from the World Bank database (http://data.worldbank.org) and includes the data from 131 countries from the year 2010 in constant 2005 USD.

From the structure of GVA we can see that these four activities form a composition, $\mathbf{x} = (x_1, x_2, x_3, x_4)'$, where x_1 corresponds to manufacturing, x_2 to agriculture, x_3 to other industry and x_4 to services. In the ternary diagram (Figure 1) proportional representations of the explanatory variables are displayed. It is visible that services forms the largest relative contribution and agriculture the smallest. These findings are reflected also in the plots of the coordinates (Figure 2), constructed according to (1), where the part of interest corresponds to x_2 (upper right), x_3 (upper left) and x_4 (lower right), respectively. Particularly, it is easy to see that the x-coordinates of the upper left and upper right plots, $z_2^{(2)}$ and $z_2^{(3)}$, are mainly negative which means that the relative contributions of agriculture and other industry are lower than the contribution of the other parts. On the other hand, the coordinate $z_2^{(4)}$ clearly shows the superiority of services.

Finally, the 3D scatterplot in Figure 2 (lower right) contains all three coordinates $z_2^{(2)}, z_3^{(2)}, z_1^{(1)}$ (in this order) to see the relation between the covariates and the response variable. Here the common dependent variable

$$z_1^{(1)} = \sqrt{\frac{3}{4}} \ln \frac{x_1}{\sqrt[3]{x_2 x_3 x_4}}$$

explains all the relative information (logratios) concerning manufacturing with respect to other three activities. The other two coordinates are formed by permutation of the remaining three parts in the denominator,

$$z_2^{(2)} = \sqrt{\frac{2}{3}} \ln \frac{x_2}{\sqrt{x_3 x_4}}; \; z_3^{(2)} = \sqrt{\frac{1}{2}} \ln \frac{x_3}{x_4}.$$

Although a certain linear relationship can be even observed from the 3D scatterplot, the orthogonal regression modeling needs to be performed in order to specify the possible influence of covariates. Accordingly, three regression models are formed, where only the explanatory coordinates $z_2^{(k)}$, $z_3^{(k)}$, $k = 2, 3, 4$

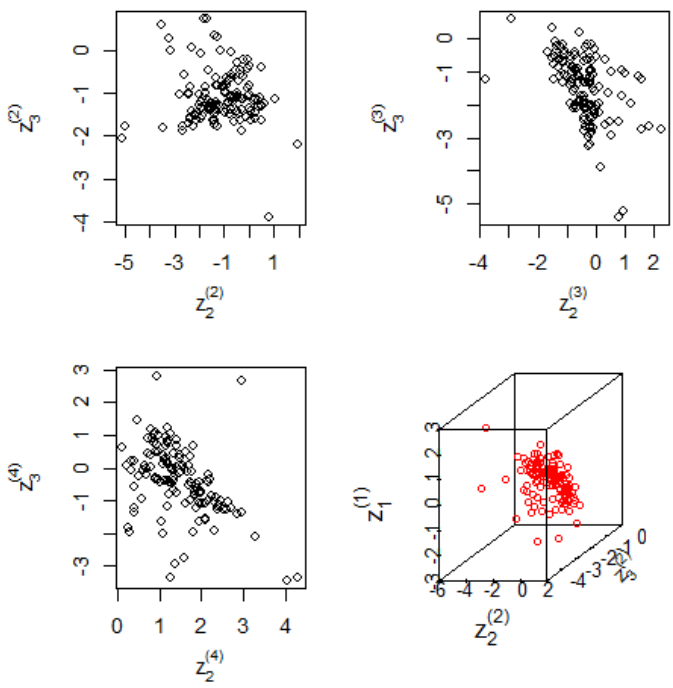


Figure 2 The plots of coordinates of explanatory variables and 3D scatterplot of coordinates $\mathbf{z}_2$.

are subject to change, while the response coordinate remains always the same. For interpretation purposes we adopt a further notation and introduce vectors $\mathbf{z}_2 = (z_2^{(2)}, z_3^{(2)}, z_1^{(1)})'$, $\mathbf{z}_3 = (z_2^{(3)}, z_3^{(3)}, z_1^{(1)})'$ and $\mathbf{z}_4 = (z_2^{(4)}, z_3^{(4)}, z_1^{(1)})'$ that follow the geometric motivation of orthogonal regression from the previous section.

parameter	$\mathbf{z}_2$	standard error	$\mathbf{z}_3$	standard error	$\mathbf{z}_4$	standard error
β_0	-2.1508366	1.0175523	-2.1508366	0.8589990	-2.1508366	0.8529446
β_1	-0.3936553	0.1472155	-0.8783889	0.7204390	1.2720442	0.6533110
β_2	-1.2415532	0.9404001	-0.9616921	0.3696728	0.2798611	0.4747517

Table 1 Estimated parameters for the orthogonal regression. The first row contains the vectors of coordinates that were used to construct the corresponding regression models.

Finally, orthogonal regression was applied in order to quantify the relations between manufacturing and GVA covariates (agriculture, other industry, services). Moreover, also the standard bootstrap methodology was employed to obtain standard errors of estimates of the regression parameters. The results from all three regression models are collected in Table 4. This table includes also the standard errors obtained from bootstrap. For interpretation purposes, the third row of the table is of particular importance as it corresponds to regression estimates (and their standard errors) of coordinates $z_2^{(2)}, z_2^{(3)}, z_2^{(4)}$ that stand for the influence of single explanatory parts.

Because different ilr coordinates represent just orthogonal rotations of each other, the absolute term parameters are the same for all regression models. Furthermore, from the values of the regression parameters in the table it is evident that the relative influence of agriculture and other industry to manufacturing (represented by the ilr coordinate $z_1^{(1)}$) is rather negative. On the other hand, from both the value of the regression parameter and the corresponding standard error we can see that services have quite strong positive influence on the relative contribution of manufacturing in the overall composition.

5 Conclusions

For a statistical analysis of compositional data it is not appropriate to apply standard statistical methods, that rely on the standard Euclidean geometry in real space, directly to the original observations. Instead, a representation of the compositions in interpretable coordinates is necessary. Since a canonical basis on the simplex, the sample space of compositional data, is not available, in this paper a particular choice of isometric logratio coordinates (1) was applied that enables to assign an own coordinate successively to each compositional part of interest. In the context of regression modeling within a composition two out of $D-1$ coordinates (1) of a D-part composition are of interest as they express the relative information concerning the response part and one of the covariates of the explanatory subcomposition. Because all coordinates contain measurement errors, the use of ordinary least squares regression for parameter estimation would lead to biased results. Therefore, orthogonal regression was used that leads also to an intuitive geometric interpretation. Although in the paper we presented just the case of three predictors, the mentioned geometric interpretation can be of course extended to the general case of $D-1$ predictors. As before, the estimated parameters $\boldsymbol{\beta}$ would be obtained using values of the normal vector formed by loadings of the smallest eigenvalue. Further inference concerning the estimates of the regression parameters, like confidence intervals and hypotheses testing, can be performed using the standard bootstrap methodology. In this contribution just the standard errors of the parameter estimates are derived.

Acknowledgements

Authors gratefully acknowledge the support of the Operational Program Education for Competitiveness - European Social Fund (project CZ.1.07/2.3.00/20.0170 of the Ministry of Education, Youth and Sports of the Czech Republic) and the grant IGA_PrF_2014_028 Mathematical Models of the Internal Grant Agency of the Palacký University in Olomouc.

References

[1] Aitchison, J.: *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London, 1986. (Reprinted in 2003 with additional material by The Blackburn Press).

[2] Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barcel-Vidal, C.: Isometric logratio transformations for compositional data analysis. *Math Geol.* **35** (2003), 279–300.

[3] Fuller, W. A.: *Measurement Error Models*. John Wiley & Sons, New York, 1987.

[4] Hron, K., Filzmoser, P. and Thompson, K.: Linear regression with compositional explanatory variables. *Journal of Applied Statistics* **39**/5 (2012), 1115–1128.

[5] Markovsky, I. and Van Huffel, S.: Overview of total least-squares methods. *Signal Processing* **87** (2007), 2283–2302.

[6] OECD: Value added, in *National Accounts at a Glance 2013*. OECD Publishing, 2013.

The application of sovereign bond spreads: The case of France, Germany and Great Britain

Jana Hvozdenska[1], Veronika Kajurova[2]

Abstract. The yield curve – specifically the spread between long term and short term interest rates is a valuable forecasting tool. It is simple to use and significantly outperforms other financial and macroeconomic indicators in predicting recessions two to six quarters ahead. The steepness of the yield curve should be an excellent indicator of a possible future economic activity. A rise in the short rate tends to flatten the yield curve as well as to slow down real growth the near term. This paper aims to analyze the dependence between slope of the yield curve and an economic activity of France, Germany and Great Britain between the years 2000 and 2013. The slope of the yield curve can be measured as the yield spread between sovereign 10-year bonds and sovereign 3-month bonds. The natural and probably the most popular measure of economic growth is by GDP growth, taken quarterly. The results showed that the best predictive lag is a lag of five quarters. The theory says that it should be lag of four quarters. The results presented also confirm that 10-year and 3-month yield spread has significant predictive power for real GDP growth after financial crisis. These findings can be beneficial for investors and provide further evidence of the potential usefulness of the yield curve spreads as indicators of the future economic activity.

Keywords: GDP prediction, yield curve, slope, spread

JEL Classification: E43, E44, E47, G01
AMS Classification: 62M20

1 Introduction

The yield curve simply plots the yield of the bond against its time to maturity. Many market observers carefully track the yield curve's shape, which is typically upward sloping and convex. However when the yield curve becomes flat or slopes downward (the spread between sovereign 10-year and 3-month bond is negative) it may signal GDP decrease (recession).

The yield curve – specifically the spread between long term and short term interest rates is a valuable forecasting tool. It is simple to use and significantly outperforms other financial and macroeconomic indicators in predicting recessions two to six quarters ahead.

This paper builds on a wide range of previous researches, but differs in some ways. Bernard and Gerlach [2] in their paper showed empirically on eight countries that the slope of the yield curve is a good predictor of the real economic activity. Berk and van Bergeijk [1] examined 12 euro-area countries over the period of 1970-1998 and found that the term spread contained only limited information about future output growth. Their work is based on the previous theoretical researches of Estrella and Hardouvelis [6], Estrella and Mishkin [7]. In their work the evidence that the slope of the yield curve and the future GDP activity are related together was proved. However it is necessary to say that this rule was true until the end of 20th century and it mostly disappeared at the beginning of 21st century and appeared again during the financial crisis (from 2008) and later on ([5], [8] [4]). Most of the studies are focused on the relationship of the yield curve and GDP activity of United States of America.

The aim of this paper is to show if the yield spread possesses the predictive power of future economic activity in France, Germany and United Kingdom and to examine which time lag of the spread is the best for prediction of the future GDP.

Despite various researches, there is not any comprehensive theory that would prove the correlation between the yield spread and economic development of the country yet. Often we come across the statements that have only theoretical basis without generally valid empirical evidence. Economic models are largely based on the argument that the yield curve tends to be flatter in the situation of the tight monetary policy and the economic slowdown typically occurs with a slight time lag.

[1] Masaryk University, Department of Finance, Lipova 41a, Brno 602 00, 174974@mail.muni.cz.
[2] Masaryk University, Department of Finance, Lipova 41a, Brno 602 00, vkajurova@seznam.cz.

Almost perfect tool containing the relevant future data provides the yield spread of government bonds. The simplest interpretation of the yield spread is through monetary policy of the country. Based on this criterion - relatively low spread reflects the restrictive and tight monetary policy and vice versa - high spread reflects loose monetary policy. We can find the theoretical justification for using of the spread in expectations hypothesis. It assumes that long term rate of return is the average of the current and expected future short term yields. The investor's decision to invest in short term or long term asset is completely irrelevant [9].

Dependence of the yield spread and GDP can be derived from their connection to the monetary policy of the state. As bond yields react to monetary policy as well as monetary policy is able to respond to the output of the economy, the yield curve assumes overlapping of policy measures and responses. The yield curve had the ability to reflect future production either directly or indirectly. Indirectly it comes to predicting of the future interest rate and the future monetary policy. It may also reflect the future production directly because the 10-year yields may depend on estimates of the output of the economy in 10-years.

A question arises – how many months, quarters, years of future economic activity can be predicted by the yield spread? Based on the study of [3] as well as [4] spread has the greatest ability in predicting one-year horizon (four quarters ahead). These studies were carried out on USA data. As it was mentioned above, to prove if the spread has the best predictive power in one-year horizon is one of the aims of this paper.

2 Methodology and data

There are many ways of using the yield curve to predict the future real activity. One common method uses inversions (when short term rates are higher than long term rates) as recession indicators. Obtaining predictions from the yield curve requires much preliminary work. There is the principle which needs to be hold: keep the process as simple as possible.

A yield curve may be flat, up-sloping, down-sloping or humped. The standard solution uses a spread (difference between two rates). The problem is to choose the spread between the right terms. The most used spread is between 10-year and 3-month bonds. The problem is that there are rarely bonds which mature exactly in 10 years (or 3 months). In that case the best solution is to use the yield curve, which shows the yield of each maturity. Creating and calculating of the yield curve is a rather difficult task because there are many ways how to do it and every country uses different model of construction.

The yield curves are constructed by Bloomberg, therefore the data for spreads were gained from Bloomberg. For the spreads were chosen 10-year government bond rates minus 3-month sovereign bond rates, as in [6] and [7]. Quarterly data were used for the spreads because the data for the economic activity are taken on quarterly basis as well. The data for real GDP can be found at Eurostat, OECD statistics or Bloomberg. The data of real GDP obtained and used in this paper are from OECD statistics.

The selected countries are France, Germany and Great Britain.

There is no previous research which would prove or reject the hypothesis of real GDP and bond spread dependence in European countries.

As a measure of real growth four-quarter percent change in real GDP was used (thus the percent change of the quarter against the last year's same quarter was calculated, e.g. the change from 1Q2004 and 1Q2003 real GDP was used). GDP is standard measure of aggregate economic activity and the four-quarter horizon answers the frequently asked question – what happens the next year?

The sample period starts from 1Q2000 and ends on 4Q2013. This time range covers the period before financial crisis, period of financial crisis and period after financial crisis. The basic model is designed to predict real GDP growth/decrease two to six quarters into the future based on the current yield spread [3].

This was accomplished by running of a series of regressions using real GDP activity and the spread between 10-year and 3-month bond yields lagged two to six quarters (e.g. if the spread was lagged by 4 quarters, the interest rate spread used for 3Q2001 is actually from 3Q2000).

The last step is to find out which spread lag is the best for which country and to prove the hypothesis that the lag of four quarters is the best one for prediction of future GDP growth.

To generate the GDP predictions the regression using the whole sample was run, and later on two divided samples of real GDP and spreads of each selected country (the sample is divided in 4Q2007/1Q2008, because this period preceded financial crisis and should show some changes in prediction of the yield curve spread) were run.

The coefficients α and β were estimated for each country:

$$Real\ GDP_{t+n} = \propto + \beta * spread_t + \varepsilon_t$$

Where:

$Real\ GDPt + n$ is a prediction of the future real GDP in time $t + n$

n is the lag of spread, value of the lag can be 2, 3, 4, 5 or 6

$spread_t$is spread between 10-year and 3-month state bonds in time t

ε_t is a white noise

3 Results and discussion

Does the yield curve accurately predict the future GDP?

To generate the GDP predictions a regression using the whole sample to generate each predicted data point was run.

3.1 Results of regression – whole sample

The whole sample of dataset contains the real GDP from 1Q2000 to 4Q2013. A regression of the whole sample was run and we got the results as seen in Table 1.

For all the countries (France, Germany and United Kingdom) the best results were gained with lag of spreads by five quarters.

We can say that models for France and Germany are statistically significant, because the p-values are under 1%(***) or 5% (**). However the R^2 are not very high. These models could be used as predictive models, but they would not explain very big proportion of the sample. The R^2 coefficients (coefficients of determination) show us how many percent of the sample can be explained by these models.

The model for United Kingdom cannot be used as a predictive due to its high p-value (0.4947) and very low R^2 (0.8677%).

Whole sample	**Constant**	**Spread**	**P – value (F – test)**	**R^2**
France (n=5)	0.00311810	0.538473	0.0421 **	0.074285
Germany (n=5)	-0.0048629	1.14207	0.0049 ***	0.137774
United Kingdom (n=5)	0.0115628	0.303703	0.4947	0.008677

Table 1 Results of all countries and whole sample from OLS regression

For example we can say that future real GDP of Germany will be:

$$\text{Real GDP Germany}_{t+5} = -0.0048629 + 1.14207 * \text{spread}_{\text{Germany t}}$$

By this model we can predict future real gross domestic product for Germany five quarters ahead.

We can test the hypothesis that the behavior of the spread and gross domestic product has changed during the financial crisis, therefore the sample was divided into two samples in order to prove this hypothesis.

3.2 Results of regression – divided samples

The research continued as follows – the whole sample was divided into two samples. The first one is from 1Q2000 to 4Q2007, the second one is from 1Q2008 to 4Q2013 in order to show if there is any change of behavior and dependency between the variables before or after the financial crisis. Regressions of the first sample and the second sample were run. The results for the time span of 1Q2000 – 4Q2007 (first sample) are possible to see in Table 2, the results for the period of 1Q2008 – 4Q2013 (second sample) are in Table 3.

1Q00 – 4Q07	Constant	Spread	P – value (F – test)	R^2
France (n=3)	0.0132012	0.514563	0.0595 *	0.113404
Germany (n=2)	0.0246873	-0.599216	0.1611	0.064403
United Kingdom (n=2)	0.0262708	0.309048	0.2575	0.042516

Table 2 Results of all countries and sample from 1Q2000 to 4Q2007

In the first period the best results for the France were gained with lag of spreads by three quarters, for Germany and United Kingdom by two quarters.

We can say that only model for France can be used as predictive, because its p-value is under 10% (*). Again the R^2 is not very high. Models for Germany and United Kingdom cannot be used as a predictive model because of their high p-values and very low R^2.

1Q08 – 4Q13	Constant	Spread	P – value (F – test)	R^2
France (n=5)	-0.0222730	1.16950	0.0003 ***	0.457461
Germany (n=5)	-0.0338772	2.37700	0.0002 ***	0.465950
United Kingdom (n=5)	-0.0350058	1.58461	0.0177 **	0.229996

Table 3 Results of all countries and sample from 1Q2008 to 4Q2013

In the second period the best results for all countries were gained by lag of spreads by five quarters. All models can be used as predictive and we got the best results from all observations. R^2 of France and Germany are higher than 40% which is very good compared to the previous period. Even the result of United Kingdom shows the big difference against the previous period. This change in prediction possibility may be caused by different behavior of financial markets after the financial crisis (after year 2008).

The best predictive models are as follows:

$$\text{Real GDP}_{\text{France t+5}} = -0.022273 + 1.1695 * \text{spread}_{\text{France t}}$$

$$\text{Real GDP}_{\text{Germany t+5}} = -0.0338772 + 2.377 * \text{spread}_{\text{France t}}$$

$$\text{Real GDP}_{\text{United Kingdom t+5}} = -0.0350058 + 1.58461 * \text{spread}_{\text{United KIngdom t}}$$

For example if there would be a change of 1% up in the spread of the France then the GDP would increase about 1.147% ($-0.022273 + 1.1695 * 1\%$).

At the end we can summarize the findings and predict the future GDP of the selected countries.

3.3 Prediction of real GDP – France, Germany and Great Britain

The spreads are known from the year 2013. The predictions of future GDP are in Table 4.

The GDP of the France should increase in all five quarters approximately about 2%. GDP of Germany should decrease about 0.13 – 0.28% in the first two quarters of 2014 and later increase approximately about 0.7% in each period. GDP of United Kingdom should decrease in first three quarters of 2014 and after that increase about 0.2% in each period.

Prediction of GDP	France		Germany		United Kingdom	
	spread	GDP	spread	GDP	spread	GDP
1Q2014	0.020099	0.00123278	0.01368	-0.0013598	0.01485	-0.01147434
2Q2014	0.020112	0.00124798	0.01306	-0.0028336	0.01354	-0.01355018
3Q2014	0.023177	0.0048325	0.01728	0.00719736	0.02037	-0.00272729
4Q2014	0.022663	0.00423138	0.01784	0.00852848	0.02321	0.001772998
1Q2015	0.02091	0.00218124	0.01683	0.00612771	0.02358	0.002359304

Table 4 Prediction of real GDP in the selected countries

At the end we can summarize the new theoretical finding according to which lag of spread is the best for predicting of the future GDP. We proved that in the selected countries it is lag of five quarters which shows the best results which are possible to use for predicting of future GDP.

At the end we can write down a short summary of the best results (n for whole sample, 1st period, 2nd period).

France – n=5, n=3, n=5

Germany – n=5, n=2, n=5

United Kingdom – n=5, n=2, n=5

The most common lag (and possibly the most suitable for predicting of the GDP) is lag of 5 quarters.

4 Conclusions

Does the yield curve accurately predict the real economic growth? Answering this seemingly simple question requires a surprising amount of preliminary work. The 10-year, 3-month spread has substantial predictive power and should provide good forecast of real growth two to six quarters into the future. We showed that the best predictive lag of spreads is a lag of five quarters in order to get the best results for predictive models. The results presented above confirm that 10-year and 3-month yield spread has a significant predictive power for real GDP growth and the behavior of the models changed during and after the financial crisis.

The simple yield curve growth forecast should not serve as a replacement for the predictions of companies, who deal with predicting of many economic indicators, it however does provide enough information to serve as a useful check on the more sophisticated forecasts.

Future research could be extended to a wider examination of the best lags of spreads in more countries, such as EU15 or EU28. It would be interesting to see if there is any rule which would prove the hypothesis that lag of five quarters is the best for predicting future GDP growth in the countries of European Union (in USA it was proved that the best lag of spread is a lag of 4 quarters).

Acknowledgements

Support of Masaryk University within the project MUNI/A/0786/2013 – *"Analysis and Prediction of Financial and Investment Products Performance"* is gratefully acknowledged.

References

[1] Berk J. and P. van Bergeijik. On the Information Content of the Yield Curve: Lessons for the Eurosystem? *Kredit und Kapital.* 2001, 28 – 47.

[2] Bernard, H. J. and S. Gerlach, Does the Term Structure Predict Recessions? The International Evidence, in *CEPR Discussion Papers 1892*, 1998.

[3] Bonser-Neal, C. a T. R. Morley. Does the Yield Spread Predict Real Economic Activity? A Multicountry Analysis, in *Economic Review*: Federal Reserve Bank of Kansas City. No. 3, 1997.

[4] Chinn, M. and K. Kucko. The predictive power of the yield curve across countries and time. *NBER Working Paper Series*, 16398, 2010.

[5] De Pace, P. GDP Growth Predictions through the Yield Spread: Time – Variation and structural Breaks, in Euro Area Business Cycle Network, 2011.

[6] Estrella, A. and G. A. Hardouvelis. The Term Structure as a Predictor of Real Economic Activity, in *Journal of Finance*, Vol. 46, No. 2, 1991, 555-576.

[7] Estrella, A. and F. S. Mishkin. The Yield Curve as a Predictor of U.S. Recessions, in Current Issues in *Economics and Finance*, Federal Reserve Bank of New York, Vol. 2, No. 7, 1996.

[8] Giacomini, R. and B. Rossi. How stable is the Forecasting Performance of the Yield Curve for Output Growth. Duke University and University of California, Los Angeles, 2005.

[9] Mishkin, F. Yield curve, in *NBER Working Papers Series*, No. 3550, 1990.

Some methodological limitations of business cycle analysis within the European area

Ladislava Issever Grochová[1], Petr Rozmahel[2]

Abstract. The paper addresses selected methodological issues of the business cycle analysis. As the choice of detrending techniques significantly influences the analysis of the classical and growth business cycles, the impact of various techniques as well as the suitability of detrending filters in context of European area convergence analysis is examined. The paper evaluates characteristics of four selected filtering techniques widely used in the literature using current macroeconomic data of the EU countries. The appropriateness of the filters is examined according to the criteria of an ideal filter defined in the paper using the periodogram analysis. The main findings indicate the Christiano-Fitzgerald filter to fit the best to the ideal filter criteria.

Keywords: Baxter-King filter, business cycle, Butterworth filter, detrending, Hodrick-Prescott filter, Christiano-Fitzgerald filter, periodogram.

JEL Classification: C44, C82, E32
AMS Classification: 90C15, 91G70, 91B84

1 Introduction

The business cycle correlation and its development over time belong among the most often evaluated characteristics when assessing the convergence and integration processes in Europe. Particularly, the correlation of business cycles is frequently used supplementary measure to evaluate the preparedness of acceding countries to join the Eurozone. The business cycle similarity was defined as a crucial criterion in a frame of the New Optimum Currency Areas Theory (OCA) as suggested by Tavlas [18] and Mongelli [14] who followed the traditional approach proposed by Mundel [15]. The cyclical component of the macroeconomic time series is estimated to identify the business cycle and measure the correlation in line with the New OCA theory approach. In addition, the dissection of the cyclical and trend component is used when estimating the economy's potential output [7] or equilibrium exchange rate such as in [16]. Despite its large theoretical foundations in the OCA theory, there is no methodological consensus in a way how to identify exactly the business cycles and measure their correlation. In our paper we focus on the problem of business cycle identification using macroeconomic time series. We follow the growth cycle approach, which considers the business cycles as fluctuations of the cyclical component of the time series around its trend component [13]. On the contrary, the classical approach considers the cycles as fluctuations of the time series in absolute levels [2]. Following the growth cycle approach, various filtering techniques are applied to estimate the cyclical component of the time series. Regarding the lack of methodological consensus there is a large discussion on the assessment of existing filters and also on estimation of an ideal filter in the business cycle literature e.g. [4], [5], [17], [6], [19], [3], [8].

The paper evaluates characteristics of four selected filtering techniques widely used in the literature using current macroeconomic data of the EU countries. The main goal is to evaluate their appropriateness according to the criteria of an ideal filter defined in the paper. The evaluation of the filters is based on the periodogram analysis. The paper is structured as follows: The next section describes used methodology. Particularly the four selected filters and an ideal filter are formally described in that section. The main results and findings are presented in the third section. The final section concludes.

2 Methodology

In this contribution, the business-cycle component of natural logarithm of the industrial production index is estimated using various filtering techniques aiming at filter appropriateness assessment. In particular, the evaluation of high-pass filtering techniques (one-parameter Hodrick-Prescott filter and two-parameter Butterworth filter) and band-pass filtering techniques (Baxter-King and Christiano-Fitzgerald filters) in condition of selected European countries in the time span 1993m1-2013m12 will be done. Because of limited availability of data the following countries are chosen: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germa-

[1] Mendel University in Brno, Department of Economics, Zemědělská 1, 613 00 Brno, ladislava.grochova@mendelu.cz.
[2] Mendel University in Brno, Department of Economics, Zemědělská 1, 613 00 Brno, rozi@mendelu.cz.

ny, Greece, Hungary, Ireland, Italy, Luxembourg, Netherlands, Poland, Portugal, Slovakia, Spain, Sweden and United Kingdom.

The time-series filters decompose a time-series into trend and cyclical components:

$$y_t = \tau_t + c_t \tag{1}$$

where τ_t is the trend component and c_t is the cyclical component. The cyclical component c_t is considered as an approximation of the growth business cycle from the OCA theory perspective. The cyclical component can be written as:

$$c_t = B(L)y_t \tag{2}$$

The ideal filter, *B(L)*, is defined as:

$$B(L) = \sum_{j=-n}^{n} B_j y_{t-j}, \; L^j y_t \equiv y_{t-j} \tag{3}$$

where B_j are the impulse–response coefficients of an ideal filter computed dependently on the filter used. The frequency domain of band-pass filters can be then specified for monthly data as:

$$B(\omega) = B(e^{-i\omega}) = \begin{cases} 1 \text{ for } \omega \in [\frac{2\pi}{96}, \frac{2\pi}{18}] \vee [\frac{-2\pi}{18}, \frac{-2\pi}{96}] \\ 0, \text{otherwise} \end{cases} \tag{4}$$

and of high-pass filters as:

$$B(\omega) = B(e^{-i\omega}) = \begin{cases} 1 \text{for } \omega \in [\frac{2\pi}{96}, \pi] \vee [-\pi, \frac{-2\pi}{96}] \\ 0, \text{otherwise} \end{cases} \tag{5}$$

i being imaginary number. Filtering techniques then approximate c_t by $\hat{c}_t$. More detailed description of the approximation using various filtering techniques is explained in the following subsections.

2.1 Hodrick-Prescott filter

The Hodrick-Prescott filtering technique (HP filter) filters data by removing the trend. The smoothness of the trend depends on a parameter λ, the trend becomes smoother as $\lambda \to \infty$. Hodrick and Prescott [10] recommended setting λ to 14,400 for monthly data. King and Rebelo [12] demonstrate that removing a trend by the HP filter corresponds to a high-pass filter as it filters low-frequencies and passes through high-frequency components in c_t. The trend component is computed solving:

$$\min_{\tau_t} \left(\sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}) \right]^2 \right) \tag{6}$$

The HP filter in a high-pass transformation can be written, in time domain, as:

$$B^{HPhigh}(L) = \frac{\lambda (1-L)^2 (1-L^{-1})^2}{1 + \lambda (1-L)^2 (1-L^{-1})^2} \tag{7}$$

Differentiating eq. (7) with respect to τ_t, cyclical component can be then expressed as two-sided moving average of y_t:

$$\hat{c}_t = B^{HPhigh}(L) y_t \tag{8}$$

where λ is a smoothing parameter (set to 14,400 for monthly data) and *L* is the lag operator.

2.2 Butterworth filter

The Butterworth (BW) filter is based on backward and forward smoothing operation of the time series filtered. The high-pass BW filter, in time domain, can be expressed as:

$$B_n^{BW}(L) = \frac{\lambda(1-L)^n(1-L^{-1})^n}{(1+L)^n(1+L^{-1})^n + \lambda(1-L)^n(1-L^{-1})^n} \tag{9}$$

where L is lag operator, $n \in N$, $\lambda = \left\{\tan(\frac{\omega_c}{2})\right\}^{-2n}$, and ω_c is the cutoff point (set to $\frac{2\pi}{96}$ for monthly data). The approximation of high-frequency cyclical component, $\hat{c}_t$, is calculated as:

$$\hat{c}_t = B_n^{BW}(L) y_t \tag{10}$$

2.3 Baxter-King filter

Baxter and King [1] divide the time-series into 3 components: trend, cycle, and irregular fluctuations. As we are interested in monthly data, business cycle for this data is according to Burns and Mitchell [2] defined as periodical components ($f_y(\omega)$) frequencies lying between 1.5 and 8 years per cycle, i.e. 18 and 96 months per cycle, with K=36. Frequencies longer than 96 months per cycle corresponding to a trend and shorter frequencies than 18 months per cycle considered to be the irregular fluctuations are then suppressed.

To isolate the business cycle component from remainder components, a two-sided finite-order (K) moving average is performed:

$$\hat{c}_t = \sum_{h=-K}^{K} B_h^{BK} y_{t-h}, K+1 \le t \le T-K \tag{11}$$

The longer moving average, the better approximation is but the higher amount of observations is to be dropped. Baxter and King find that three years of observations (i.e. K=36 months) should be dropped to make the filter perform well.

Weights obtained from the minimization problem in the frequency domain are:

$$\min_{\hat{B}_j^{BK},\, j=-K,\ldots,K} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|B(e^{-i\omega}) - \hat{B}^{BK;K,K}(e^{-i\omega})\right|^2 d\omega \tag{12}$$

2.4 Christiano-Fitzgerald filter

The Christiano-Fitzgerald random walk (CF) filter is constructed on the same principles as the BK filter formulating detrending and smoothing problem in the frequency domain [11]. In contrast to the BK filter, no observations must be dropped by the CF filter. The cyclical component is approximated as follows:

$$\hat{c}_t = \sum_{h=-K}^{K} \hat{B}_j^{CFp,f} y_{t-j} \tag{13}$$

where $f = T-t$, $p = t-1$, $\hat{B}_j^{p,f}$ being a solution of minimization problem in the frequency domain.

$$\min_{\hat{B}_j^{CFp,f},\, j=-f,\ldots,p} = \int_{-\pi}^{\pi}\left|B(e^{-i\omega}) - \hat{B}^{CFp,f}(e^{-i\omega})\right|^2 f_y(\omega) d\omega \tag{14}$$

$f_y(\omega)$ being the spectral density function of y_t.

2.5 Filters' appropriateness assessment

A well-performing filter is able to extract the periodical component c_t of a time series of desired frequencies. An ideal filter passes or blocks the stochastic cycles at given frequencies by having a gain of 1 or 0. For monthly data, high-pass filters pass through the stochastic cycles from a specified frequency $\omega_0 = \frac{2\pi}{96}$ for monthly data and block lower frequencies. Band-pass filters allow stochastic cycles in the specified range of frequencies $\left[\omega_0 = \frac{2\pi}{96}, \omega_1 = \frac{2\pi}{18}\right]$ corresponding to a cycle length of 1.5–8 years expressed in months to pass through and block all the other frequencies. The goodness of filters performance can be evaluated using periodogram [9].

The evidence of how well the cyclical component c_t was estimated, is presented by periodogram. A periodogram is an estimator of the spectral density function of y_t, $f_y(\omega)$. A periodogram identifies dominant frequencies in an observed time series. A relatively large value of periodogram indicates relatively more importance for the respective frequency in explaining the periodicity in the observed time series. That is, if applied to the filtered series, only desired frequencies should appear to be dominant. If the filter completely drops the unwanted frequencies, i.e. it behaves ideally, the log periodogram is a flat line at the minimum value of -6 for unwanted frequencies. The log periodogram values greater than -6 at unwanted frequencies indicate a limited ability of the filter to block undesired frequencies outside the specified band $\left[\omega_0 = \frac{2\pi}{96}, \omega_1 = \frac{2\pi}{18}\right]$.

3 Results and discussion

In the next section, the assessment of the filters appropriateness in the European conditions based on periodogram analysis is done.

First, we explore whether the filters pass only desired frequencies, while those unwanted attenuate. This is shown in periodograms. The periodogram plots a filter estimate of the business-cycle component displaying the results in natural frequencies, which are the standard frequencies divided by 2π. If the filter performs well, log periodogram values equal -6 for undesired frequencies, whereas greater values than -6 (but not greater than 6) are expected for desired frequencies. The frequencies that are supposed to pass differ according to the filter used.

When the HP and BW filters are converted into high-pass filters, the lowest frequency they should pass is $\omega_0 = \frac{2\pi}{96}$ for monthly data. This implies that lower frequencies should be blocked and their respective log periodogram values should equal -6. In contrast higher frequencies should pass, even those that represent irregular fluctuation of high frequencies. The cutoffs for high-pass filters are represented in the plot (see Fig. 1) by a vertical line showing the natural-frequency cutoff (at 1/96 = 0.01042). The HP filter, in general, passed more undesired frequencies, $\omega \in (0, \frac{2\pi}{96})$, than BW filter. This can be seen namely in the case of Belgium, Estonia, Ireland, Portugal or Slovakia. In these countries the HP filter unlike the BW filter erroneously passed frequencies below the frequency cutoff showing the log periodogram values greater than -6. From this point of view the BW filter performs better in European conditions than the HP filter.

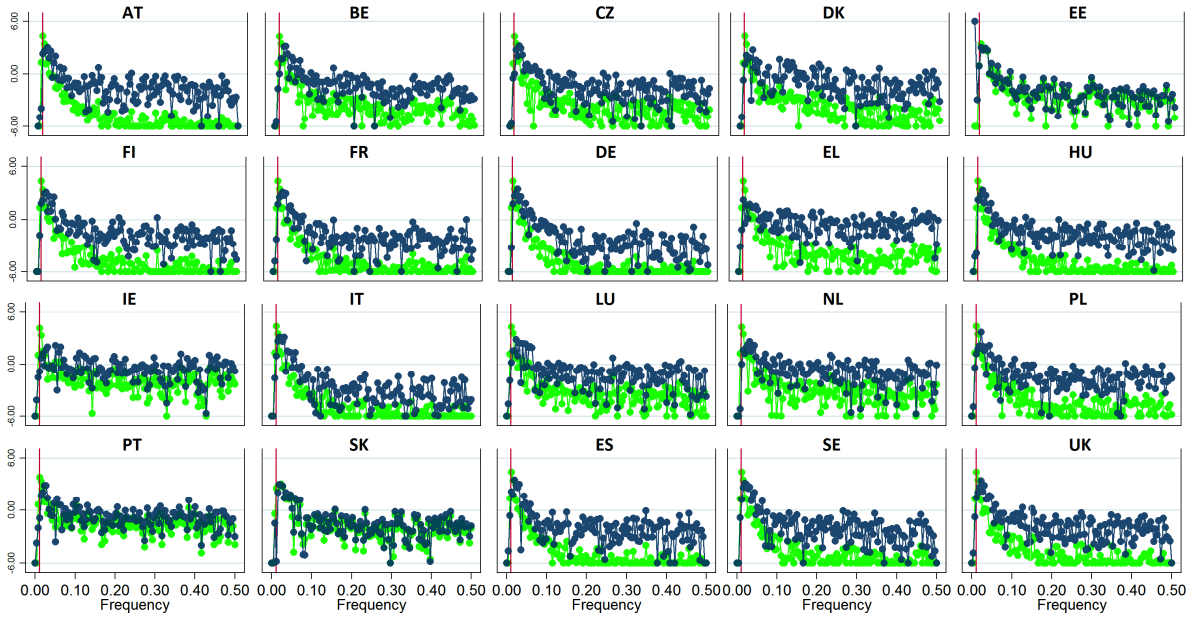


Figure 1 Periodogram of logarithms of industrial production filtered by the HP (blue) and BW (green) filters in the selected EU countries

Band pass filters should be able to block all frequencies outside the range $\left(\frac{2\pi}{96}, \frac{2\pi}{18}\right)$ that corresponds to a cycle length of 1.5–8 years extracting so both trend and irregular fluctuation components. The vertical lines in the plot (see Fig. 2) show the lower natural-frequency cutoff (at 1/96 = 0.01042) and the upper natural-frequency cutoff (at 1/18 = 0.05556) that limit the interval of desired frequencies. An ideal filter should imply log periodogram values greater than -6 only in the specified band. Fig. 2, however, presents higher log periodogram values and so a king of inability of band pass filters to attenuate undesired, mainly higher, frequencies. This is remarkable in case of Austria, Belgium, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Netherlands, Spain and Sweden for BK filter and Denmark, Estonia, Finland, Hungary, Ireland, Poland, Portugal, Slovakia, Spain and United Kingdom for CF filter. Almost in all cases (except for Denmark, Poland, Portugal, Slovakia and United Kingdom) the CF filter eliminates better unwanted frequencies when compared to the BK filter.

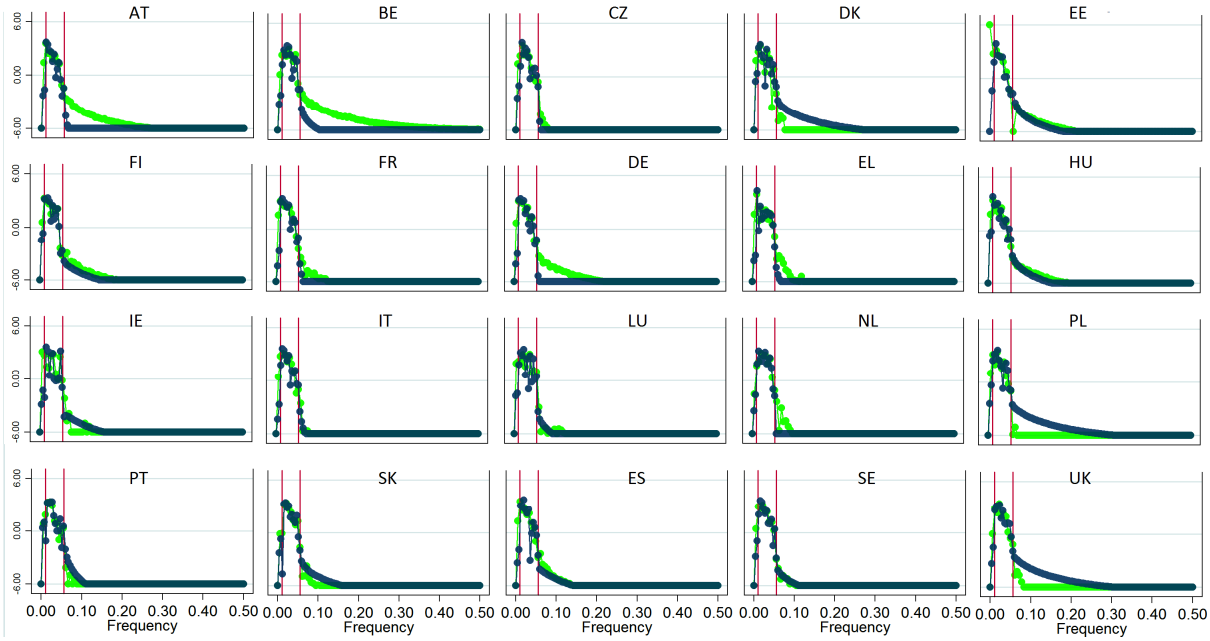


Figure 2 Periodogram of logarithms of industrial production filtered by the CF (blue) and BK (green) filters in the selected EU countries

4 Conclusions

Using the periodogram analysis we evaluated the appropriateness of four detrending filters from the perspective of an ideal filter using current macroeconomic data of 20 selected EU economies. From the perspective of an ideal filter in terms of passing desired and blocking unwanted frequencies it seems that CF filter performs the best. It satisfyingly suppresses frequencies greater that the higher cutoff when compared to the BK filter and as unlike the HP filter or BW filter it eliminates irregular fluctuations. In particular, except for 5 out of 20 European countries the results of CF filter are the most convincing since the filter eliminates majority of unwanted frequencies. Using this application framework these findings represent another kind of support to the CF filter besides that of methodological advantages provided by Christiano and Fitzgerald [11].

Acknowledgements

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 290647.

References

[1] Baxter, M., and King. R. G.: Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics* **81** (1999), 575–593.
[2] Burns, A. F., and Mitchell. W. C.: *Measuring Business Cycles*. New York: National Bureau of Economic Research, 1946.
[3] Buss, G.: Preliminary Results on Asymmetris Baxter-King Filter. *Journal of Applied Mathematics* **4** (2011), 3, 239–248.
[4] Canova, F.: Detrending and Business Cycle Facts. *Journal of Monetary Economics* **41** (1998), 533–540.
[5] Canova, F.: Does Detrending Matter for the Determination of the Reference Cycle and the Selection of Turning points? *The Economic Journal* **452** (1999), 109, 126–150.
[6] Estrella, A.: Extracting Business cycle Fluctuations: What Do Time Series filters Really Do? *Federal Resereve Bank of New York Staff Report* **289** (2007), 34 p.
[7] Garrat, A., James, M. and Shaun, P. Measuring output gap nowcast uncertainty. *International Journal of Forecasting* **30** (2014), 268-279.
[8] Grochová, L. and Rozmahel, P. Business cycle correlation of the CEEC and the Euro area: some methodological controversy. In: *Proceedings of 30th International Conference Mathematical Methods in Economics* (Ramík, J. and Stavárek, D., eds.). Czech Econometric Society, 2012, 237-242.
[9] Hamilton, J. D.: *Time Series Analysis*. Princeton: Princeton University Press, 1994.
[10] Hodrick, R. J., and Prescott. E. C.: Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking* **29** (1997), 1–16.
[11] Christiano, L. J., and Fitzgerald. T. J.: The band pass filter. *International Economic Review* **44** (2003), 435–465.
[12] King, R. G., and Rebelo. S. T.: Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control* **17** (1993), 207–231.
[13] Lucas, R.E.: Understanding Business Cycles. In: *Stabilisation Domestic and International Economy* (Brunner, K., and Meltzer, A.H., eds.). Carnegie-Rochester Conference Series on Public Policy, 1977, 7–29.
[14] Mongelli, P. F.: New Views on The Optimum Currency Area Theory: What is EMU Telling Us? *Working Paper ECB* **138**, 2002.
[15] Mundell, R. A Theory of Optimum Currency Areas. *The American Economic Review* **51** (1961), 657-665.
[16] Plecitá, K. and Střelec, L. Behavioral equilibrium exchange rate in Greece and Ireland. In: *Proceedings of 30th International Conference Mathematical Methods in Economics* (Ramík, J. and Stavárek, D., eds.). Czech Econometric Society, 2012, 715-720.
[17] Schenk-Hoppé, K. R.: Economic Growth and Business Cycles: A Critical Comment on Detrending Time Series. University of Zurich, Institute for Empirical Research in Economics, *Working Paper* **54**, 2001.
[18] Tavlas, G. The New Theory of Optimum currency Areas. *The World Economy* **16** (1993), 663-685.
[19] Watson, M.: How Accurate Are Real-Time Estimates of Output Trends and Gaps? *Economic Quarterly* **93** (2007), 2, 143–161.

Alternative approach of assessing European countries efficiency from the perspective of sustainable economic performance

Ladislava Issever Grochová[1], Jitka Janová[2], David Hampel[3], Luboš Střelec[4]

Abstract. Nowadays, the goal of European economies, that is often emphasized (e.g. Europe 2020 Strategy), is the focus on both economic performance and sustainability. The economic performance is often tightly accompanied by pollution and wastefulness as negative by-products of economic performance that burden the life quality. GNP or GDP as such are from the point of view of capturing well-being inadequate metrics. Particularly, in its economic-social and environmental dimensions, some of the aspects that are often referred to as sustainability GNP or GDP are inconvenient measures. For this reason it is necessary to take both economic and environmental aspects into account. Alternative indicators are often used for such a measure of sustainable economic performance. We offer an alternative or even complementary approach based on Data Envelopment Analysis (DEA) to identify and quantify the multifactor impacts of economic growth on social development, and to measure and evaluate the comparative effectiveness of different countries in terms of sustainable economic development.

Keywords: Data Envelopment Analysis, EU28, sustainable economic performance.

JEL Classification: C44, D61, Q01, Q32
AMS Classification: 90B50, 90C05

1 Introduction

Recently, an attitude to an environment has dramatically changed since limited amount of resources was scientifically confirmed and the threat of their depletion emerged. It is then clear that environmental issues became an important aspect of economic performance. While economic performance itself is measured as values of goods and services exchanged in the market, majority of environmental goods and services do not have this value as they are not marketed. This can be regarded as a typical market failure leading to overexploitation and degradation of these resources.

In order to deal with this market failure lacking measures of environmental goods and services values has been recognized as an important item of information that should be supplied. Thus, the construction of Environmental Performance Indices (EPIs) has become the aim of most of the environmentally directed studies. According to [11], the approaches to developing EPIs can be divided into two main categories – direct and indirect approach. The majority of contributions in this field represent the indirect approach [3]. They focus on the construction of overall indices aggregating weighted key indicators. DEA represents a direct approach. It is a well-developed nonparametric technique for evaluating relative efficiencies of decision making units (DMUs) with multiple inputs and outputs [7]. Nonetheless, the adoption of DEA in the context of environmental performance measurements is still scarce, e. g. [5], [6], [10], [11] or [12]. These studies discuss various technical parameters and debate characteristics of environmental DEA technologies. Nonetheless, environmental DEA applications are still needed.

To enrich the less developed field offering an alternative evaluation of economies in terms of sustainable development, in this contribution we follow the direct approach to environmental-economic-social performance evaluation – Data Envelopment Analysis (DEA). There are various advantages of this technique. First, a multi-

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Economics, Zemědělská 1, Brno, 613 00, ladislava.grochova@mendelu.cz.
[2] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, 613 00, jitka.janova@mendelu.cz.
[3] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, 613 00, qqhampel@mendelu.cz.
[4] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, 613 00, lubos.strelec@mendelu.cz.

dimensional perspective or independence on individual preferences can be mentioned. In contrast to Environmental performance indicators (EPIs), DEA can elaborate multiple inputs and outputs without any requirement on functional relationships between inputs and outputs or market value assessment. Finally, it is a technique measuring the relative efficiency of decision making units (DMUs) showing which inputs and outputs cause the inefficiency of a DMU and to what extent in some of particular DEA versions. For the analysis Stata software is used [9].

The paper aims at alternative evaluation of sustainable development using DEA, in particular, BCC model. The rest of the paper is organized as follows: section 2 introduces methodology and description of DEA, section 3 dissects results and the last section concludes.

2 Methodology

In an economy inputs (x) are consumed in order to produce desirable outputs (y). The production possibility set then represents all technically feasible combinations of inputs and outputs:

$$\Psi(x)=\{(x,y)|x \text{ can produce } y, x\in R_a^+, y\in R_b^+\}. \tag{1}$$

Input requirements can be defined as

$$I(y)=\{x\in R_a^+|(x,y)\in\Psi\}. \tag{2}$$

Strong disposability and no free lunch axioms, respectively, are assumed [8]:

$$\forall(x,y)\in\Psi, \text{ if } x'\geq x \text{ and } y'\leq y \text{ then } (x',y')\in\Psi \tag{3}$$

$$\forall(x,y)\notin\Psi \text{ if } x=0, y\geq 0, y\neq 0. \tag{4}$$

For assesment of countries efficiency we use the additive DEA model, see [4]. Namely we employ BCC model, introduced by Banker, Charnes and Cooper [2], which measures technical efficiency as the convexity constraint ensures that the composite unit is of similar scale size as the unit being measured. The resulting efficiency is always at least equal to the one given by the CCR model, and those DMUs with the lowest input or highest output levels are rated efficient. Unlike the CCR model, the BCC model allows for variable returns to scale.

DEA either maximizes output subjected to a given amount of inputs or minimizes the amount of inputs for a given output. As the set of goods and services produces with a given amount of inputs is often accompanied with socially undesired resource depletion and pollution, DEA in ecological framework must be carefully applied when dealing with the environmental impact. DEA, in this context, is supposed to maximize the economic goods and services output but to consider negative environmental effects at the same time.

According to Allen [1] there are four possibilities how to cope with pollutants:

1. converting pollutants by taking their reciprocal or by subtracting the pollutant from a maximal value regarding the converted pollutant as usual output,
2. regarding the pollutant as an input that should be, together with other inputs, minimized,
3. regarding pollutants as the only type of input, factors being subtracted from the value of a product,
4. regarding pollutants indirectly subtracting them from the product.

As negative environmental effects of production should be minimized, we follow the second possibility how to treat pollutants. Then we can add undesirable pollutant (u) as an input into the equation (1) as:

$$\Psi(x)=\{(x,u,y)|(x,u) \text{ can produce } y, x\in R_a^+, y\in R_b^+, u\in R_c^+\}. \tag{5}$$

Formally, we evaluate the relative efficiency of the specific oth country as follows:

$$\begin{aligned} &\max es^- + es^+ \\ &\text{subject to}: X\lambda + s^- = x_O, \\ &Y\lambda - s^+ = y_O, \\ &e\lambda = 1, \end{aligned} \tag{6}$$

$$\lambda \geq 0,\ s^- \geq 0,\ s^+ \geq 0.$$

Here, n is the number of decision making units, k is the number of inputs, m is the number of outputs, $X = (x_j)$ is $k \times n$ matrix of inputs, $Y = (y_j)$ is $m \times n$ matrix of outputs, e is a row vector with all elements equal to 1, s^- is a vector of input slacks, s^+ is a vector of output slacks, x_O is a column vector of inputs of the oth decision making unit, and $\lambda \in R^n$ is a vector connecting inputs and outputs.

For the evaluation of particular countries performance we use country-level data in 2004, 2008 and 2012. This allows us to make a comparison of efficiencies providing so a kind of dynamic analysis. In order to assess the efficiency in sustainable development, 3 inputs and 1 output are used taking into account usage of data in this field and limited data availability.

In particular, we deal with carbon dioxide emissions in millions of kilotons (CO2), energy use in kg of oil equivalent (Energy) both representing ecological dimension, and risk of poverty rate measured as 60% of median equalized income after social transfers (Poverty) being a proxy of social dimension. These quantities are used as inputs to produce current GDP in millions of US dollars (GDP), i. e. economic dimension.

Presented data generally provide proxies of environmental, economic and social dimensions, the aspects of sustainable development. Descriptive statistics of the data for different EU subsets are shown in Table 1 presented below.

EU28					
Variable	Obs	Mean	Std. Dev.	Min	Max
CO2	84	0.389	0.311	0.097	1.870
Energy	84	143.262	42.498	79.164	263.929
Poverty	84	16.004	3.800	9.000	25.900
GDP	84	586219	899454	5600	3600000
Large-size countries					
Variable	Obs	Mean	Std. Dev.	Min	Max
CO2	15	0.430	0.347	0.097	1.430
Energy	15	170.656	53.199	88.658	258.115
Poverty	15	14.527	2.722	11.000	20.200
GDP	15	1721600	178656	560000	2400000
Small-size countries					
Variable	Obs	Mean	Std. Dev.	Min	Max
CO2	69	0.381	0.305	0.129	1.870
Energy	69	137.370	37.686	79.164	263.929
Poverty	69	16.325	3.938	9.000	25.900
GDP	69	676232	966884	14000	3600000
CEEC					
Variable	Obs	Mean	Std. Dev.	Min	Max
CO2	24	0.490	0.432	0.138	1.870
Energy	24	134.408	53.348	79.164	263.929
Poverty	24	16.092	3.835	9.000	22.200
GDP	24	486850	780746	5600	2700000
V4 countries					
Variable	Obs	Mean	Std. Dev.	Min	Max
CO2	12	0.520	0.338	0.194	1.220
Energy	12	136.377	45.241	88.658	223.211

Poverty	12	15.283	3.771	9.000	20.500
GDP	12	773617	1041223	5600	2700000
Periphery countries					
Variable	Obs	Mean	Std. Dev.	Min	Max
CO2	12	0.458	0.279	0.210	1.270
Energy	12	164.308	36.630	118.518	237.141
Poverty	12	16.517	4.133	11.000	23.400
GDP	12	145667	77591	41000	270000

Table 1 Descriptive statistics of the data according to different EU subsets
Note: Countries: V4: Czech Republic, Hungary, Poland, Slovakia; CEEC: Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia; Periphery: Greece, Ireland, Portugal, Spain. Large-size: Germany, France, Italy, Poland, Spain, United Kingdom. Small-size: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Portugal, Romania, Slovakia, Slovenia, Sweden.

3 Results and discussion

The results of the DEA are presented in Table 2. Table 2 contains the DEA results for all dimensions, i.e. environmental, economic and social dimension – sustainable development aspects, as well as the results for economic and environmental dimension, and economic and social dimension, separately, for all explored time periods: 2004, 2008 and 2012.

As far as all sustainable aspects together are concerned, the best performing countries are the so-called old EU members, except for the Czech Republic in all 3 years observed and Malta in 2004. All these countries lie on the production frontier which means that the given amount of inputs produced maximally attainable output. In contrast, the inefficient countries are, overall, the CEE countries. In terms of economic-environmental issues the most inefficient DMUs are the CEE countries plus Sweden and Finland, while the most efficient DMUs are western and southern countries. In particular, the best performance from economic and environmental dimension is attributed to Germany, the United Kingdom and Ireland. The socio-economic dimension leaders are Germany, the Czech Republic and the Netherlands, while the worst performers are located in the CEE region.

Comparing particular dimensions, there is a country that is efficient in all these 3 aspects and that is Germany. All other countries are either efficient in some dimensions only or ale directly inefficient. The representatives of the first group are the United Kingdom, the Czech Republic, France or Ireland. The UK and Ireland are efficient in overall sustainable development aspects and environmental dimension; they perform worse in social tasks. Whereas the Czech Republic and France do well in the overall issue and social issues contrary to environmental dimension. Among overall inefficient countries can be mentioned Bulgaria, Romania or the Baltic states.

	all dimensions (economic, environmental and social)			**economic and environmental dimensions**			**economic and social dimensions**		
Country	**2004**	**2008**	**2012**	**2004**	**2008**	**2012**	**2004**	**2008**	**2012**
AUT	0.908	0.943	0.995	0.783	0.772	0.737	0.821	0.746	0.804
BEL	0.759	0.770	0.797	0.537	0.530	0.497	0.738	0.638	0.671
BGR	0.703	0.524	0.530	0.355	0.392	0.364	0.693	0.421	0.441
CYP	0.842	0.800	0.878	0.764	0.692	0.701	0.646	0.566	0.662
CZE	1.000	1.000	1.000	0.420	0.481	0.461	1.000	1.000	1.000
DEU	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DNK	1.000	1.000	1.000	0.851	0.834	0.804	0.961	0.775	0.768
ESP	0.873	0.869	0.939	0.841	0.858	0.848	0.546	0.522	0.535
EST	0.531	0.557	0.607	0.363	0.395	0.345	0.515	0.462	0.560
FIN	0.971	0.810	0.838	0.395	0.428	0.394	0.950	0.666	0.720
FRA	0.928	1.000	1.000	0.765	0.742	0.746	0.872	1.000	1.000
GBR	1.000	1.000	1.000	1.000	1.000	1.000	0.623	0.661	0.834

GRC	0.879	0.827	0.750	0.837	0.827	0.750	0.526	0.455	0.463
HRV	0.761	0.790	0.690	0.693	0.717	0.654	0.578	0.520	0.460
HUN	0.799	0.853	0.827	0.590	0.577	0.546	0.770	0.726	0.710
IRL	1.000	1.000	1.000	1.000	1.000	1.000	0.500	0.583	0.646
ITA	0.990	0.954	0.961	0.983	0.930	0.880	0.601	0.632	0.664
LTU	0.532	0.619	0.703	0.449	0.527	0.588	0.507	0.450	0.510
LUX	0.853	0.874	0.906	0.660	0.713	0.664	0.819	0.672	0.721
LVA	0.648	0.657	0.688	0.573	0.657	0.553	0.536	0.347	0.516
MLT	1.000	0.983	0.941	0.924	0.958	0.887	0.727	0.588	0.628
NLD	1.000	1.000	1.000	0.695	0.696	0.658	1.000	0.940	0.981
POL	0.600	0.689	0.716	0.524	0.554	0.552	0.511	0.557	0.581
PRT	0.857	0.825	0.822	0.811	0.814	0.785	0.512	0.488	0.546
ROU	0.602	0.588	0.615	0.471	0.536	0.523	0.578	0.385	0.441
SVK	0.797	0.901	0.838	0.417	0.511	0.514	0.782	0.826	0.754
SVN	0.878	0.872	0.846	0.593	0.604	0.564	0.852	0.732	0.721
SWE	1.000	1.000	1.000	0.519	0.550	0.543	0.934	0.767	0.737
CEEC	0.709	0.726	0.737	0.475	0.523	0.501	0.675	0.590	0.623
Core	0.937	0.954	0.966	0.808	0.822	0.799	0.826	0.765	0.809
Periphery	0.935	0.913	0.912	0.915	0.904	0.869	0.543	0.548	0.577
V4	0.799	0.861	0.845	0.488	0.531	0.518	0.766	0.777	0.761

Table 2 DEA results of sustainable economic efficiencies within selected EU countries

The evolution of efficiencies in particular dimension can be observed in the Fig. 1. Regarding the sustainable development the worst performers are the CEE countries, in which a mild improvement during the period 2004–2012 can be seen. However, among these countries a group of V4 countries is considerably more efficient. The old EU members perform continuously well. A significant difference right between the old and new members is visible in environmental issues. Here the old members perform by almost 100% better than the new members, however with a slight decrease in the last period. In the social dimension, the Core and V4 countries are the most efficient with a stable performance in the time period analyzed. Nevertheless, the less efficient countries improve moderately their performance.

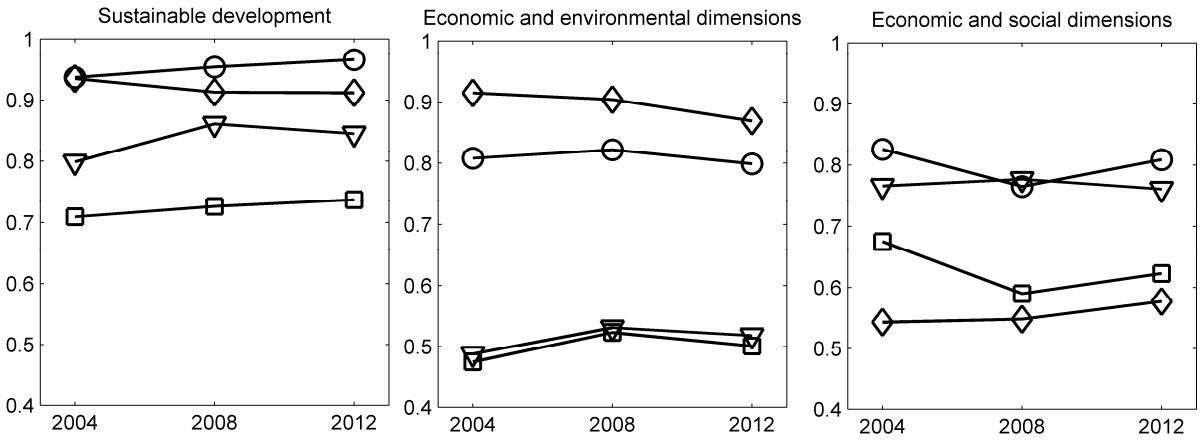


Figure 1 Efficiency in the selected EU countries (CEEC square, Core countries[5] circle, Periphery countries diamond, V4 countries triangle) during years 2004–2012

[5] **Core**: Austria, Belgium, Denmark, Finland, France, Germany, Italy, Luxembourg, Netherlands, Sweden, United Kingdom.

4 Conclusions

We show possibilities of the DEA approach in the area of countries description and comparison. At first, calculated efficiencies allow to sort countries and evaluate correctness of established frontiers. Variables from different insights form different results; therefore it is possible to identify lacks of particular countries. Efficiencies calculated for data from different years enable assessing of dynamical characteristics, as improving or deteriorating tendency of a country. Moreover, the same as for countries holds for groups of countries, where average efficiency brings meaningful information.

To sum up, better performers are the old EU members that are efficient in either all or some issues. The new members are less efficient mainly in environmental dimension. In contrast, the evolution of efficiencies is mainly in favour of the new EU members.

Acknowledgements

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 290647. We also thank to Prof. Kabát, whose insightful comments helped us to improve the paper considerably.

References

[1] Allen, K.: Data Envelopment Analysis in the Service Sector. Harzer wirtschaftswissenschaftliche Schriften, 1999, 203–235.
[2] Banker, R. D., Charnes, A., Cooper, W. W.: Some models for estimating technical and scale efficiencies in data envelopment analysis. *Management Science* **30** (1984), 1078–1092.
[3] Bell, S., Morse, S.: *Sustainability Indicators: Measuring the Immeasurable?* Earthscan, 2008.
[4] Charnes, A., Cooper, W. W., Golany, B., Seiford, L.: Foundations of data envelopement analysis for Pareto Koopmans efficient empirical production functions, *Journal of Econometrics*, **30** (1985), 91–107.
[5] Färe, R, Grosskopf, S.: Modeling undesirable factors in efficiency evaluation: comment. *European Journal of Operational Research* **157** (2004), 242–245.
[6] Färe, R, Grosskopf, S., Tyteca, D.: An activity analysis model of the environmental performance of firms application to fossil-fuel-fired electric utilities. *Ecological Economics* **18** (1996), 161–175.
[7] Ramanathan, R.: An Introduction to Data Envelopment Analysis: a tool for performance measurement. Sage Publications, New Delhi, 2003.
[8] Shephard, R.W.: Theory of Cost and Production Function, Princeton University Press, Princeton (NJ), 1970.
[9] StataCorp.: Stata Statistical Software: Release 13. College Station, TX: StataCorp LP, 2013.
[10] Zaim, O.: Measuring Environmental Performance of State Manufacturing through Changes in Pollution Intensities: A DEA Framework. *Ecological Economics* **48** (2004), 37–47.
[11] Zhou, P., Ang, B. W., Poh, K. L.: Measuring environmental performance under different environmental DEA technologies. *Energy Economics* **30** (2008), 1–14.
[12] Zofio, J. L., Prieto, A. M.: Environmental Efficiency and Regulatory Standards: the Case of CO2 Emissions from OECD Industries. *Resource and Energy Economics* **23** (2001), 63–83.

Multiplicative efficiency and cross-efficiency: a comparative study

Josef Jablonský[1]

Abstract. The aim of the paper is to compare conventional DEA models with multiplicative DEA models which is a non-traditional class of models for evaluation of efficiency. Multiplicative efficiency of a unit under evaluation is defined as a maximum ratio of Cobb-Douglas type function of outputs divided by the same type of function of inputs. The paper discusses how the multiplicative models can be used for efficiency evaluation and for ranking of the set of units. Selected super-efficiency multiplicative models and multiplicative cross-efficiency evaluation is used in the study. Two super-efficiency concepts – Andersen and Petersen models and Tone's SBM model – are applied. They are compared to the results given by cross-efficiency model that was introduced in 1986 as a tool for ranking of efficient units in DEA. In this concept the units are evaluated using the optimal weights of other efficient units. The results given by conventional approach and the new one are compared on an example with a real economic background.

Keywords: data envelopment analysis, efficiency, cross-efficiency, super-efficiency, Cobb-Douglas.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Data envelopment analysis (DEA) is a non-parametric technique for evaluation of relative efficiency of decision making units described by multiple inputs and outputs. Let us suppose that the set of decision making units (DMUs) contains n elements. The DMUs are evaluated by m inputs and r outputs with input and output values x_{ij}, $i = 1,2,\ldots,m$, $j = 1,2,\ldots,n$ and y_{kj}, $k = 1,2,\ldots,r$, $j = 1,2,\ldots,n$, respectively. The efficiency of the q-th DMU is expressed in conventional DEA models as the weighted sum of outputs divided by the weighted sum of inputs with weights reflecting the importance of single inputs/outputs v_i, $i = 1,2,\ldots,m$ and u_k, $k = 1,2,\ldots,r$, under the assumption that the efficiency of all other units in the set is limited to unity. This efficiency score of the unit under evaluation can be found by solving the following optimization problem:

Maximize

$$\theta_q = \frac{\sum_{k=1}^{r} u_k y_{kq}}{\sum_{i=1}^{m} v_i x_{iq}}$$

subject to (1)

$$\frac{\sum_{k=1}^{r} u_k y_{kj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \qquad j = 1,2,\ldots,n,$$

$$u_k \ge \varepsilon, \qquad k = 1, 2,\ldots, r,$$
$$v_j \ge \varepsilon, \qquad i = 1, 2,\ldots, m,$$

where $\varepsilon > 0$ is a non-Archimedean infinitesimal that ensures positive weights u_k and v_i. The problem (1) is not linear in its objective function but can be easily transformed into a linear program – for more details see e.g. (Charnes et al., 1978) or (Zhu, 2002) .

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, 130 67 Praha 3, Czech Republic, e-mail: jablon@vse.cz.

The paper compares conventional DEA models and multiplicative DEA models that are based on definitions of virtual inputs and outputs using Cobb-Douglas type functions. Main attention in this comparison is paid to models for ranking of DMUs and among them to models based on cross evaluation and super-efficiency models. Next section of the paper contains basic definitions and formulates multiplicative DEA models and models for ranking of efficient units. Section 3 compares presented models on the real data set of 20 DMUs (Czech banks). Final part of the paper discusses advantages and disadvantages of presented approaches and future research directions.

2 Multiplicative DEA models

Let us suppose that the inputs x_{ij} and outputs y_{kj} of all DMUs are all greater than unity. Then the multiplicative virtual outputs and inputs of the j-th DMU can be expressed as:

$$\prod_{i=1}^{m} x_{ij}^{v_i} \text{ and } \prod_{k=1}^{r} y_{kj}^{\mu_k}\ , j = 1, 2,\ldots, n, \tag{2}$$

where v_i, $i = 1, 2,\ldots, m$, and μ_k , $k = 1, 2,\ldots, m$, are virtual input and output multipliers that have to be greater or equal than unity. Charnes et al. (1982) define, according to the conventional DEA model (1), the following multiplicative DEA model:

Maximize

$$\psi_q = \frac{\prod_{k=1}^{r} y_{kq}^{\mu_k}}{\prod_{i=1}^{m} x_{iq}^{v_i}}$$

subject to (3)

$$\frac{\prod_{k=1}^{r} y_{kj}^{\mu_k}}{\prod_{i=1}^{m} x_{ij}^{v_i}} \le 1, \qquad j = 1,2,...,n,$$

$$u_k \ge 1, \qquad k = 1, 2,\ldots, r,$$
$$v_j \ge 1, \qquad i = 1, 2,\ldots, m.$$

Model (3) can be re-written as standard linear programming problem using logarithms to any base. The new model is very close to conventional DEA SBM model. Its primal (4) and dual (5) forms are as follows:

maximize

$$\log_p(\psi_q) = \sum_{k=1}^{r} \mu_k y'_{kq} - \sum_{i=1}^{m} v_i x'_{iq},$$

subject to (4)

$$\sum_{k=1}^{r} \mu_k y'_{kj} - \sum_{i=1}^{m} v_i x'_{ij} \le 0, \quad j = 1,2,...,n,$$
$$\mu_k, v_i \ge 1, \qquad k = 1,2,...,r, i = 1,2,...,m,$$

and

minimize

$$\log_p(\psi_q) = -\left(\sum_{k=1}^{r} s_i^- + \sum_{i=1}^{m} s_k^+ \right),$$

subject to (5)

$$\sum_{j=1}^{n} x'_{ij} \lambda_j + s_i^- = x'_{iq}, \qquad i = 1,2,...,m,$$

$$\sum_{j=1}^{n} y'_{kj}\lambda_j - s_k^+ = y'_{kq}, \qquad k = 1, 2,...,r,$$
$$\lambda_j \geq 0, \qquad j = 1, 2,\ldots,n,$$
$$s_i^- \geq 0, s_k^+ \geq 0, \qquad i = 1, 2,..., m, k = 1, 2,...,r,$$

where $y'_{kj} = \log_p(y_{kj})$ and $x'_{ij} = \log_p(x_{ij})$ with any p. Models (4) and (5) are not unit invariant, i.e. multiplication of any input or output by a constant influences the efficiency score of the units. This problem was solved in (Charnes et al., 1983) where an invariant multiplicative model was derived. It is proved that adding one identical dummy input and output to the set of inputs and outputs leads to unit invariant multiplicative model.

Ranking of DMUs using DEA models is widely discussed task. Inefficient units can be easily ranked according to their efficiency scores, the efficient ones have maximum efficiency score – e.g. unity in models (1) or (3) – and this fact does not allow its direct ranking. There are available various models that allow ranking of efficient units in DEA models. One of the first one is the super-efficiency model by Andersen and Petersen (1993). Since 1993 variety of models based of quite different concepts were proposed. Some of the most important ones are summarized in (Jablonsky, 2012). In this paper we discuss the application of the following three modelling concepts (they are described in (Jablonsky, 2012) in more detail):

1. Andersen and Petersen model that is based on removal of the unit under evaluation from the set of units. By this removal the efficiency score of the unit is greater that unity and the units can be ranked easily according to this measure (super-efficiency score).
2. SBM super-efficiency model introduced by Tone (2002). Its multiplicative (logarithmic) formulation follows:

   minimize

$$\psi_q^{SBM} = \frac{\frac{1}{m}\sum_{i=1}^{m} x_i^* / x'_{iq}}{\frac{1}{r}\sum_{k=1}^{r} y_k^* / y'_{kq}},$$

   subject to (6)

$$\sum_{j=1}^{n} x'_{ij}\lambda_j + s_i^- = x'_{iq}, \qquad i = 1,2,...,m,$$
$$\sum_{j=1}^{n} y'_{ij}\lambda_j - s_k^+ = y'_{kq}, \qquad k = 1,2,...,r,$$
$$x_i^* \geq x'_{iq}, \qquad i = 1,2,...,m,$$
$$y_k^* \leq y'_{kq}, \qquad k = 1,2,...,r,$$
$$\lambda_j \geq 0, \qquad j = 1,2,\ldots,n, j \neq q,$$
$$\lambda_q = 0.$$

   The optimal objective function value of model (6) is greater than 1 for efficient DMUs – higher value is assigned to more efficient units. Model (6) is not linear but can easily be transformed into a linear program and solved using common linear solvers.
3. Multiplicative cross-efficiency evaluation is another concept which allows ranking of DMUs in this class of models. The basic idea of this concept is evaluation of each DMU using optimal multipliers (weights) of inputs and outputs given by model (3). Let us denote μ_{kj}, $k = 1, 2, \ldots, r$, $j = 1, 2, \ldots, n$ optimal multiplier of the k-th output given by the model (3) in evaluation of the unit j-th DMU and v_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$ optimal multiplier of the i-th input given by the same model. Cross efficiency of the q-th DMU by using optimal weights of the j-th DMU is defined as follows:

$$E_{qj} = \frac{\prod_{k=1}^{r} y_{kq}^{\mu_{kj}}}{\prod_{i=1}^{m} x_{iq}^{v_{ij}}}, \quad q = 1,2,...,n,\ j = 1,2,...n. \tag{7}$$

It is clear that $E_{qq} = \psi_q$, i.e. by using optimal weights of the unit DMU_q its cross efficiency is equal to the efficiency score given by model (3). Geometric average cross efficiency φ_q is defined as follows:

$$\varphi_q = \left[\prod_{j=1}^{n} E_{qj} \right]^{1/n}, \ q = 1,2,\ldots,n. \tag{8}$$

It is clear that $\varphi_q \in (0,1>$ and that the upper bound can appear in very special cases only. That is why the values φ_q can be used for complete ranking of all DMUs.

3 Numerical comparison

Results of the above formulated models are illustrated in this section on efficiency evaluation of 20 banks operating on the Czech financial market. It is rather an illustrative example than a serious case study even though the data set used in the example has real background. The data set is taken from public financial statements of the banks for year 2010. The following inputs and outputs are considered (all characteristics for 20 banks are presented in Table 1):

Inputs:
- Equity of the bank in millions of CZK (Czech crowns),
- Number of full time employees (FTE).

Outputs:
- Deposits in millions of CZK,
- Credits in millions of CZK.
- Profit in millions of CZK.

Bank	Equity	FTE	Credits	Deposits	Profit
CS	14 014	10 163	416 854	530 101	14 317
Citibank	673	900	125 062	169 425	14 310
ČSOB	3 635	6 420	209 172	568 199	13 572
GE Money	775	2 290	97 262	109 942	3 851
Hypotecní Bank	157	476	145 070	455	2 288
ING CZ	84	123	19 169	89 211	821
KB	6 556	7 883	334 834	441 285	14 417
LBBW Bank	727	350	19 161	18 929	50
Raiffeisenbank	989	528	152 663	125 936	2 320
UniCredit	1 578	265	172 070	174 373	3 473
Volksbank	186	726	39 147	30 155	345
Wustenrot	380	200	31 978	35 956	341
Commerzbank	42	371	40 162	19 950	343
Czech Export B	39	144	59 856	58 690	217
Czech-Moravian	171	218	20 776	27 076	1 053
PPF Bank	44	154	18 655	36 332	804
Blue pyramid	399	356	49 030	69 119	1 026
Raiffeisen SS	56	257	39 578	76 160	830
CS SS	420	250	44 307	97 540	1 464
Wustenrot hyp	17	33	11 261	2 259	90

Table 1 Data set of 20 Czech banks

Table 2 contains experimental results (efficiency scores) based on the given data set completed by the rank of the units according to the results of the model:

(1) First column of the Table 2 contains efficiency scores computed using CCR input oriented model derived from model (1). The values lower than unity indicate inefficiency. The efficient units have normally maximum efficiency score 1 but here is the value greater than 1 (so called super-efficiency score) computed using Andersen and Petersen model (mentioned above). The best performers among the DMUs according to this model are Citibank, UniCredit and Czech Export Bank.

(2) For comparison purposes with the CCR model the second column presents efficiency score of the unit invariant multiplicative (logarithmic) model (4) or (5) – it is the model extended by one identical dummy input and output (MULTe). This model identifies more units as efficient and does not allow their direct ranking. It is interesting that several units with very low CCR efficiency scores are efficient in this model (e.g. CS and KB bank)

(3) Third column contains efficiency (lower than 1) and super-efficiency scores (greater than 1) calculated using SBM and super SBM model (Tone, 2002). As in the first case Citibank, UniCredit and Czech Export Bank belong among the best units.

(4) Multiplicative SBM (MSBM) model (6) is one of the possibilities how to discriminate among efficient units in MULTe model. Higher values are assigned to more efficient units.

(5) Cross efficiency evaluation based on the original data set with two inputs and three outputs. It is clear that ranking of banks according to this kind of evaluation is closely correlated to the one given by conventional CCR model.

(6) Multiplicative cross (MCross) evaluation based on the unit invariant model, i.e. model with one dummy input and output. In both cross evaluations the problem with potential alternative optimal solution is not taken into account. This problem is discussed in detail in (Doyle and Green, 1994) for conventional DEA models and in (Cook and Zhu, 2014) for multiplicative DEA models. The results in Table 2 show that the tanking given by MCross model copies more or less the ranking given by conventional cross evaluation.

Bank	**CCR-I**	**MULTe**	**SBM**	**MSBM**	**Cross**	**MCross**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
CS	0.1373 (19)	1.0000 (1)	0.1099 (20)	1.0057 (10)	0.0905 (19)	0.0025 (17)
Citibank	2.3542 (1)	1.0000 (1)	2.1046 (1)	1.1183 (2)	0.7282 (3)	0.5853 (1)
ČSOB	0.2295 (16)	1.0000 (1)	0.2182 (17)	1.0163 (9)	0.1697 (15)	0.0159 (10)
GE Money	0.2716 (15)	0.0464 (17)	0.2228 (16)	0.8819 (18)	0.1629 (16)	0.0083 (12)
Hypotecní Bank	1.0812 (6)	1.0000 (1)	1.0406 (6)	1.0246 (7)	0.4119 (9)	0.0036 (16)
ING CZ	1.8061 (4)	1.0000 (1)	1.4568 (4)	1.0499 (5)	0.7911 (2)	0.1286 (5)
KB	0.1724 (18)	1.0000 (1)	0.1528 (19)	1.0029 (11)	0.1281 (17)	0.0080 (13)
LBBW Bank	0.1117 (20)	0.0003 (20)	1.0000 (8)	1.0000 (12)	0.0521 (20)	0.0000 (20)
Raiffeisenbank	0.6578 (10)	0.3866 (12)	0.5169 (12)	0.9724 (14)	0.3687 (10)	0.0565 (7)
UniCredit	2.0270 (2)	1.0000 (1)	1.5135 (3)	1.0578 (4)	0.5203 (6)	0.1806 (4)
Volksbank	0.1786 (17)	0.0104 (19)	0.1628 (18)	0.8209 (20)	0.1238 (18)	0.0010 (19)
Wustenrot	0.3522 (14)	0.0263 (18)	0.2659 (15)	0.8434 (19)	0.2038 (14)	0.0011 (18)
Commerzbank	0.7897 (9)	0.1405 (14)	0.6158 (10)	0.9442 (15)	0.2743 (13)	0.0292 (8)
Czech Export B	2.0022 (3)	1.0000 (1)	1.9820 (2)	1.0755 (3)	0.8868 (1)	0.5077 (2)
Czech-Moravian	0.4200 (12)	0.0799 (16)	0.3841 (13)	1.0000 (12)	0.3156 (11)	0.0063 (14)
PPF Bank	1.0152 (7)	1.0000 (1)	1.0076 (7)	1.0438 (6)	0.5959 (5)	0.1115 (6)
Blue pyramid	0.4182 (13)	0.0879 (15)	0.3325 (14)	0.8927 (17)	0.2863 (12)	0.0099 (11)
Raiffeisen SS	1.1941 (5)	1.0000 (1)	1.0971 (5)	1.0164 (8)	0.6356 (4)	0.3845 (3)
CS SS	0.6530 (11)	0.2671 (13)	0.5259 (11)	0.9327 (16)	0.4216 (8)	0.0276 (9)
Wustenrot hyp	0.8542 (8)	1.0000 (1)	1.0000 (8)	1.2381 (1)	0.4409 (7)	0.0051 (15)

Table 2 Multiplicative efficiency, cross-efficiency and super-efficiency

Comparison of results presented in Table 2 shows that Citibank and Czech Export Bank belong to the top units according to all applied models. In the contrary, the worse units are Volksbank and LBBW Bank.

4 Conclusions

The idea of multiplicative DEA models is quite old and this kind of models was almost lost for wider applications in present but in general the idea is interesting and represents many opportunities for further development. A significant advantage of conventional DEA models consists in fact that the results can be explained to decision makers quite easily and they can understand them without any problems. Explanation of results in case of multiplicative models is not so easy. It is big disadvantage of multiplicative models. Despite this the multiplicative models can find their applications in ranking of DMUs identified as efficient in one of the conventional DEA models. Further research in this field is open.

Acknowledgements

The research is supported by the Czech Science Foundation – project no. P403/12/1387.

References

[1] Andersen, P. and Petersen, N.C.: A procedure for ranking efficient units in data envelopment analysis. *Management Science*, **39** (1993), 1261-1264.

[2] Cook, W.D.and Zhu, J.: DEA Cobb-Douglas frontier and cross-efficiency. *Journal of the Operational Research* Society, **65** (2014), 265–268.

[3] Charnes, A., Cooper, W.W. and Seiford, L.: Invariant multiplicative efficiency and piecewise Cobb-Douglas envelopments. *Operations Research Letters*, **2** (1983), 101-103.

[4] Charnes, A., Cooper, W.W., Seiford, L. and Stutz, J.: A multiplicative model for efficiency analysis. *Socio-Economic Planning Sciences*, **16** (1982), 223–224.

[5] Charnes, A., Cooper, W.W., and Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2** (1978), 429–444.

[6] Doyle, J. and Green, R.: Efficiency and cross efficiency in DEA: Derivations, meanings and uses. *Journal of the Operational Research Society*, **45** (1994), 567-578.

[7] Jablonsky, J.: Multicriteria approaches for ranking of efficient units in DEA models. *Central European Journal for Operations Research*, **20** (2012), 435–449.

[8] Tone, K. A slack-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, **143** (2002), 32-41.

[9] Zhu, J.: *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*. Kluwer Academic Publishers, Boston, 2002.

Lexicographic Optimal Public Service System Design

Jaroslav Janáček[1], Marek Kvet[2]

Abstract. A public service system structure is usually formed by deployment of limited number of service centers and the associated objective is to minimize total disutility like social costs, which are proportional to the distances travelled by all users to the nearest service center. In such case, the weighted *p*-median problem with min-sum criterion is solved. In many cases such criterion leads to the situation, when the total social costs are minimal, but the disutility of the worst situated user is extremely high. This contribution deals with such public service system design, where the disutility of the worst situated users is minimized first and then disutility of better located users is optimized under the condition that the disutility of the worst situated users does not worsen. Two different approaches to this lexicographical min-max problem will be presented and compared here. The first of them makes use of the location-allocation model and the second one is based on the radial formulation. Both methods are implemented in commercial IP-solver. Various benchmarks are solved by both methods and the results are compared from the point of computational time and solution accuracy.

Keywords: Public service system design, weighted *p*-median problem, location-allocation model, radial formulation

JEL Classification: C61
AMS Classification: 90C06, 90C10, 90C27

1 Introduction

Fair designer of a public service system pays more attention to the worst situated system users than to an average user. This approach leads to various schemes, which can be applied to the design. The strongest way of the fairness is denoted as lexicographic min-max optimization [2]. When the lexicographic min-max problem is solved, then a sequential process of optimization is used. First, the system issue for the worst situated user is to be minimized, then the issue of the second worst situated user is proceed unless the optimized issue of the previously processed user be worsen. This process is to be repeated for all other levels of the possible issue.

The public service system design (PSSD) can be formulated as a problem, where at most p service centers can be deployed in a finite set of possible center locations so that a disutility perceived by the system users is minimal. In a host cases, the disutility perceived by an individual user is derived from his distance to the nearest located service center. When an average user disutility is minimized, the public service system design can be dealt as the weighted *p*-median problem [1], [3]. Attempts at applying the *p*-median problem solution on real instances of the public service system showed that the solving techniques based on the original location-allocation model often failed due to memory or computational time demands. The necessity of solving large instances of the *p*-median problem started the development of other approaches based on radial formulation of the problem [1], [3], [4]. The radial formulation of the *p*-median problem avoids the formalization of assigning a center to a user and, this way, the model is smaller than the location-allocation one. In addition, branch-and-bound method used for solving the radially formulated *p*-median problem converges much faster than the same method applied on location-allocation formulated problem. The fair design of a public service system is much more complicated problem than the problems, where the average disutility or the sum of all users' disutility values is minimized. In the case, when the issue for system users gets its value only from a finite set of values, Ogryczak's approach can be employed [7]. This approach subsequently minimizes components of so called distribution vector, where the first component corresponds to the number of users, which obtain the highest value of the issue. The second component represents the number of users, which get the second highest value of the issue, and so on.

[1] Prof. RNDr. Jaroslav Janáček, CSc., University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, e-mail: Jaroslav.Janacek@fri.uniza.sk
[2] Ing. Marek Kvet, PhD.,University of Žilina, University Science Park, Univerzitná 8215/1, 010 26 Žilina, Slovakia, e-mail: marek.kvet@uniza.sk

This paper deals with the lexicographic optimal design of the public service system. The both mentioned ways of the problem modelling, radial and location-allocation ones will be tested with the purpose to find decision, which of them is more suitable for the lexicographical optimization. The remainder of the paper is organized as follows. The second section presents the Ogryczak's approach applied on location-allocation formulation of the weighted *p*-median problem. The third section introduces the approximate approach, which uses radial formulation instead of the location-allocation one. The both approaches are compared in the computational study performed with data describing heath emergency system design for eight Slovak regions.

2 Ogryczak´s lexicographic approach to PSSD

The public service system design with minimal disutility is a task of location of at most *p* service centers so that the sum of individual disutilities of each user coming only from the nearest located service center is minimal. To describe the problem, we denote I a set of possible service center locations and J the set of possible users' locations. The symbol b_j denotes the number of users located at the location j. The disutility for a user at the location j following from the possible center location i is denoted as d_{ij}. Under assumption that the user's disutility originates only from the nearest service center, the disutility g_j for a user for a user at the location j can be described as follows:

$$g_j(\mathbf{z}) = \sum_{i \in I} d_{ij} z_{ij} \tag{1}$$

The allocation variables $z_{ij} \in \{0,1\}$ for each $i \in I$ and $j \in J$ used in (1) are introduced to assign a customer j to a possible location i by the value of one. In addition, let the decision of service center location at the place $i \in I$ be modeled by a zero-one variable $y_i \in \{0,1\}$ which takes the value of 1 if a center is located at i and it takes the value of 0 otherwise. Then the constraints of location-allocation model can be formulated as follows.

$$\sum_{i \in I} z_{ij} = 1 \quad for\ j \in J \tag{2}$$

$$z_{ij} \le y_i \quad for\ i \in I, j \in J \tag{3}$$

$$\sum_{i \in I} y_i \le p \tag{4}$$

$$z_{ij} \in \{0,1\} \quad for\ i \in I, j \in J \tag{5}$$

$$y_i \in \{0,1\} \quad for\ i \in I \tag{6}$$

In the above model, the allocation constraints (2) ensure that each customer is assigned to exactly one possible center location. Link-up constraints (3) enable to assign a customer to a possible location i only if the service center is located at this location and the constraint (4) bounds the number of located service centers.

Let the range of all disutility values be represented by a finite set of ordered values $G_0, G_1, \ldots, G_w, G_{w+1}$, so that $G_{max} = G_0 > G_1 > \ldots > G_w > G_{w+1} = G_{min}$. The set of ordered values may contain either all disutility values or the set may represent a series of points dividing the range into $w+1$ zones $(G_{u+1}, G_u]$ for $u = 0, 1, \ldots, w$.

For the first case, where $\{G_u\}$ covers set of all disutility values, let $B_u(\mathbf{z})$ be defined by (7).

$$B_u(\mathbf{z}) = \sum_{\substack{j \in J \\ g_j(\mathbf{z}) = G_u}} b_j \quad for\ u = 0, ..., w \tag{7}$$

The lexicographic min-max problem according to [7] consists in lexicographic minimizing of the vector $[B_0(\mathbf{z}), B_1(\mathbf{z}), \ldots, B_w(\mathbf{z})]$ subject to (2) – (6).

For the second case, where $\{G_u\}$ is only set of dividing points chosen from set of all general disutility values, let $B_u(\mathbf{z})$ be defined by (8).

$$B_u(\mathbf{z}) = \sum_{\substack{j \in J \\ g_j(\mathbf{z}) \in (G_{u+1}, G_u]}} b_j \quad for\ u = 0, ... w \tag{8}$$

The lexicographic min-max problem according to [7] consists then in lexicographic minimizing of the vector $[B_0(\mathbf{z}), B_1(\mathbf{z}), \ldots, B_w(\mathbf{z})]$ subject to (2) – (6).

The nonlinearity in the expressions (7) or (8) can be excluded by introducing a series of slack non-negative variables h_{tj} for given stage t and for $j \in J$. The variables h_{tj} must be connected to the system of original variables by link-up constraints (9) and obligatory constraints (10).

$$h_{uj} \geq \sum_{i \in I} d_{ij} z_{ij} - G_{u+1} \qquad for\ u = 0, \ldots t,\ \ j \in J \tag{9}$$

$$h_{uj} \geq 0 \quad for\ u = 0, \ldots t,\ \ j \in J \tag{10}$$

The constraints (11) where $\underline{B}^*_u$ denotes the optimal objective function value for $u = 0, \ldots, t\text{-}1$, ensure that optimization at the stage t does not spoil the objective function value achieved at the previous stages.

$$\sum_{j \in J} b_j h_{uj} \leq \underline{B}^*_u \quad for\ u = 0, \ldots t-1 \tag{11}$$

After the adjustments, the optimal solution of the lexicographic min-max problem can be obtained by step-by-step solving series of the following problems, where the *t-th* problem is formulated as follows.

$$Minimize \ \sum_{j \in J} b_j h_{tj} \tag{12}$$

Subject to (2) - (6) *and* (9) - (11)

3 Radial approach to the lexicographic optimal PSSD

The radial approach was originally developed for solving of *p*-median problem [1], [3], [4]. The weighted *p*-median problem is formulated as a task of location of at most *p* service centers so that the sum of individual disutility values of each user coming only from the nearest located service center is minimal. To describe the problem, we denote *I* a set of possible service center locations and *J* the set of possible users' locations as above. The symbol b_j denotes the number of users located at the location *j*. The disutility for a user at the location *j* following from the possible center location *i* is denoted as d_{ij}. Within this paper, we focus on so called radial formulation used in the above papers and we explore the approaches based on the set of dividing points. To model decision on locating center at particular location, we also use a zero-one variable $y_i \in \{0, 1\}$, which takes the value of 1, if a center should be located at the location *i*, and it takes the value of 0 otherwise. To obtain an upper or a lower bound of the original objective function, the range $[d_0, d_m]$ of all possible disutilities $d_0 < d_1 < \ldots < d_m$ from the matrix $\{d_{ij}\}$ is partitioned into $r+1$ zones. The zones are separated by a finite ascending sequence of so called *dividing points* $D_1, D_2 \ldots D_v$ chosen from the sequence $d_0 < d_1 < \ldots < d_m$, where $0 = d_0 = D_0 < D_1$ and also $D_v < D_{v+1} = d_m$. The zone *s* corresponds with the interval $(D_s, D_{s+1}]$. The length of the *s*-th interval is denoted by e_s for $s = 0 \ldots v$. In addition, auxiliary zero-one variables x_{js} for $s = 0 \ldots v$ are introduced. The variable x_{js} takes the value of 1, if the disutility of the user at $j \in J$ from the nearest located center is greater than D_s and it takes the value of 0 otherwise. Then the expression $e_0x_{j0} + e_1x_{j1} + e_2x_{j2} + \ldots + e_vx_{jv}$ constitutes an upper approximation of the disutility d_{j*} from user location *j* to the nearest located service center. If the disutility d_{j*} belongs to the interval $(D_s, D_{s+1}]$, then the value of D_{s+1} is the upper estimation of d_{j*}. Let us introduce a zero-one constant a_{ij}^s for each triple $[i, j, s] \in I \times J \times [0 \ldots v]$. The constant a_{ij}^s is equal to 1, if the disutility d_{ij} between the user at *j* and the possible center location *i* is less or equal to D_s, otherwise a_{ij}^s is equal to 0. Then the radial weighted covering model can be formulated according to [5], [6] as follows:

$$Minimize \qquad \sum_{j \in J} b_j \sum_{s=0}^{v} e_s x_{js} \tag{13}$$

$$Subject\ to: \quad x_{js} + \sum_{i \in I} a_{ij}^s y_i \geq 1 \quad for\ j \in J,\ s = 0, 1, \ldots, v \tag{14}$$

$$\sum_{i \in I} y_i \leq p \tag{15}$$

$$x_{js} \geq 0 \quad for\ j \in J,\ s = 0, 1, \ldots, v \tag{16}$$

$$y_i \in \{0,1\} \quad for\ i \in I \tag{17}$$

The objective function (13) gives the upper bound of the sum of original disutility values. The constraints (14) ensure that the variables x_{js} are allowed to take the value of 0, if there is at least one center located in radius D_s from the user location j. The constraint (15) puts a limit p on the number of located facilities.

The issue $g_j(\mathbf{x})$ of the designed system for an individual user located at a place j is estimated here by (18).

$$g_j(\mathbf{x}) = \sum_{s=0}^{v} e_s x_{js} \tag{18}$$

Now we could use the same lexicographic formulation as in the previous section and derive the model for the sequential optimization of the vector $[B_0(\boldsymbol{x}), B_1(\boldsymbol{x}), \ldots, B_w(\boldsymbol{x})]$, where the components of vector are defined accordingly to (7) or (8) for variables $\boldsymbol{x}$ instead of $\mathbf{z}$. The process could be repeated for an arbitrary sequence of the ordered values $G_0, G_1, \ldots, G_w, G_{w+1}$, so that $G_{max} = G_0 > G_1 > \ldots > G_w > G_{w+1} = G_{min}$. Nevertheless, we make use of the special structure of the dividing points where $0 = D_0 < D_1 < \ldots < D_v < D_{v+1}$ and thanks to it, we can considerably simplify the resulting model. Let us determine the sequence $\{G_u\}$ so that we set $w=v$ and $G_u = D_{v+1-u}$. Under this adjustment, the right-hand-side of the constraints (9) can be reformulated as shown in (19).

$$\sum_{s=0}^{w} e_s x_{js} - G_{u+1} = \sum_{s=0}^{v} e_s x_{js} - D_{v-u} = \sum_{s=0}^{v} e_s x_{js} - \sum_{s=0}^{v-u+1} e_s \tag{19}$$

As the variable h_{uj} was originally defined by nonlinear expression (20), what includes both series of constraints (9) and (10), we obtain the substitution constraints (21) for h_{uj}.

$$h_{uj} = \max\{0, \sum_{s=0}^{w} e_s x_{js} - G_{u+1}\} \tag{20}$$

$$h_{uj} = \sum_{s=v-u}^{v} e_s x_{js} \tag{21}$$

Using the substitution (21), the constraints (11) can be adjusted to the constraints (22).

$$\sum_{j \in J} b_j \sum_{s=v-u}^{v} e_s x_{js} \le \underline{B}_u^* \quad for\ u = 0, \ldots t-1 \tag{22}$$

Now the optimal solution of the lexicographic min-max problem can be obtained by step-by-step solving series of the following problems for $t=0, \ldots, v$.

$$Minimize \quad \sum_{j \in J} b_j \sum_{s=v-t}^{v} e_s x_{js} \tag{23}$$

Subject to (14) - (17) *and* (22).

4 Computational study

This Section deals with the lexicographic optimal design of the public service system. Two approaches to the problem are compared here. The approaches differ in the way of the problem modelling. The first of the approaches is based on the Ogryczak's approach applied on location-allocation formulation of the weighted p-median problem and the second one uses radial formulation instead of the location-allocation one. The goal of this computational study is to decide, which of them is more suitable for lexicographical optimization. The approaches were tested on the pool of benchmarks obtained from the road network of Slovak Republic. The instances are organized so that they correspond to the administrative organization of Slovakia. For each self-governing region (Bratislava - BA, Banská Bystrica - BB, Košice - KE, Nitra - NR, Prešov - PO, Trenčín - TN, Trnava - TT and Žilina - ZA) all cities and villages with corresponding number of inhabitants b_j were taken. The coefficients b_j are rounded to hundreds. The number of possible service center locations $|I|$ is the same as the number of user locations $|J|$ in all solved instances and the network distance from user to the nearest located center was taken as an individual user´s disutility. The experiments were organized so that the range of all possi-

ble disutility values for lexicographic optimization was reduced by computation of the maximal relevant disutility **DmM**. For this purpose the location-allocation model minimizing the highest relevant disutility was designed. This method is denoted by **MinMax** in all following tables. The computational time in seconds necessary for finding the optimal value of **DmM** is plotted in the column **Time**. After the determination of **DmM**, the lexicographic approach was applied to get the resulting system design. Particular model described by (12), (2)-(6), (9)-(11) was solved step-by step. For the resulting design, the objective function **ObjF** was computed in accordance to (24). The number of most impacted users whose disutility reaches the limit **DmM** is denoted by **BD**.

$$F(\mathbf{y}) = \sum_{j \in J} b_j \min\{d_{ij} : i \in I,\ y_i = 1\} \tag{24}$$

To compare the location-allocation and radial formulations used in the lexicographic optimization process, each instance was solved in both ways. In the following tables we compare both methods from the point of computational time, because the objective function values were the same. The radial formulation is denoted by **Radial** and the location-allocation approach is denoted by **Loc-Alloc.** All experiments were performed using the optimization software FICO Xpress 7.5 (64-bit, release 2013) for both location-allocation model and the covering approach with different quality criteria. The associated code was run on a PC equipped with the Intel® Core™ i7 2630QM processor with the parameters: 2.0 GHz and 8 GB RAM. The experiments were performed with the data for each self-governing region of Slovakia. The original problem comes from real emergency medical service system, where given number of ambulances should be deployed in the region so that ensure the rescue service for associated population. To obtain bigger pool of benchmarks, the value of p was set in such a way, that the ratio of $|I|$ to p equals 2, 3, 4, 5, 10, 15, 20, 30, 40, 50 and 60 respectively. The associated results for the region of Žilina containing all generated instances are plotted in the Table 1.

$\|I\|$	p	MinMax		FAIR - LEX MIN-MAX			
		Time	DmM	ObjF	BD	Radial Time	Loc-Alloc Time
315	158	16.37	4	6115	176	0.08	3.20
315	105	38.92	6	19339	1	0.53	8.22
315	79	38.89	7	19979	160	1.17	17.91
315	63	48.13	8	24083	387	1.83	24.24
315	32	48.63	14	29576	13	4.12	45.43
315	21	64.74	16	41488	79	5.87	61.35
315	16	60.95	20	55001	38	14.24	94.30
315	11	68.56	25	79464	8	37.30	227.65
315	8	85.49	31	79490	2	63.27	289.13
315	7	156.41	34	97661	18	119.14	1146.68
315	6	157.12	34	102427	81	109.67	671.84

Table 1 Results of the experiments for the self-governing region of Žilina

Since the detailed results for other regions had similar characteristic as obtained for the region of Žilina, we report only selected instances for the other regions. The value of parameter p in the selected instances was chosen so that it corresponds to the original set of problems from real emergency medical service system.

	$\|I\|$	p	MinMax		FAIR - LEX MIN-MAX			
			Time	DmM	ObjF	BD	Radial Time	Loc-Alloc Time
BA	87	22	0.96	8	18996	7	0.08	0.92
BB	515	52	450.50	13	31209	25	4.93	83.26
KE	460	31	201.48	15	44130	2	179.59	417.55
NR	350	35	112.74	13	37370	3	99.51	176.89
PO	664	45	1418.45	15	44760	1	138.86	401.80
TN	276	28	29.94	12	32120	22	23.57	79.76
TT	249	25	68.89	13	27469	5	13.04	83.45
ZA	315	32	48.63	14	29576	13	4.12	45.43

Table 2 Results of the experiments for the self-governing regions of Slovakia for selected parameter p

The last results evaluation consists of enumeration of average computational times, which were obtained for groups of instances derived from eight self-governing regions of Slovakia. The instances were grouped according to the ratio of $|I|$ to p into eleven groups.

$\|I\| / p$	MinMax	FAIR LEX MIN-MAX	
		Radial	Loc-Alloc
2	68.24	0.12	5.24
3	175.80	0.48	11.09
4	240.12	1.32	15.57
5	247.64	2.15	27.58
10	229.63	27.18	104.13
15	313.07	126.59	253.10
20	378.14	81.94	520.37
30	862.99	581.48	1381.32
40	612.93	1667.30	2863.35
50	1070.71	200.42	3628.34
60	724.50	453.73	5793.89
AVERAGE	**441.18**	**281.79**	**1223.77**

Table 3 Average results for all self-governing regions of Slovakia grouped by $|I|/p$

5 Conclusions

The main goal of this paper was to present and compare different approaches to the lexicographic optimal public service system design. The first approach was based on the Ogryczak's procedure applied on location-allocation formulation of the weighted *p*-median problem. The second approach uses the radial formulation instead of the location-allocation one. Comparing the reported results in Tables 1, 2 and 3, we can conclude that the more suitable way of the problem modelling consists in the radial formulation. It follows from the presented results that the computational time of the radial approach is in order smaller than the location-allocation approach. Thus we can conclude that we have presented useful tool for the large fair public service system design, which can be implemented using common commercial optimization software. Further research in this area could be aimed at the possibility of reducing the time necessary for obtaining the optimal value of **DmM** by using the radial formulation in the min-max model.

Acknowledgement

This work was supported by the research grants VEGA 1/0296/12 "Public Service Systems with Fair Access to Service", VEGA 1/0339/13 "Advanced microscopic modelling and complex data sources for designing spatially large public service systems", APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks" and by the following project: University Science Park of the University of Žilina (ITMS: 26220220184) supported by the Research & Development Operational Program funded by the European Regional Development Fund.

References

[1] Avella, P., Sassano, A., Vassil'ev, I..: *Computational study of large scale p-median problems*. In: Mathematical Programming 109, pp. 89-114, (2007).

[2] Bertsimas, D., Farias, V. F., Trichakis, N.: *The Price of Fairness*. Oper. Res. 59 (2011), pp. 17–31

[3] García, S., Labbé, M., Marín, A.: *Solving large p-median problems with a radius formulation*. In: INFORMS Journal on Computing 23 (4), pp. 546-556, (2011).

[4] Janáček, J.: *Approximate Covering Models of Location Problems*. In: Lecture Notes in Management Science: ICAOR ´08, Vol. 1, Sept. 2008, Yerevan, Armenia, pp. 53-61, ISSN 2008-0050, (2008).

[5] Janáček, J., Kvet, M.: *Approximate solving of large p-median problems*. In: ORP3-Operational research peripatetic post-graduate programme: proceedings, Cádiz, Spain, September 13-17, 2011, Cádiz: Servicio de Publicaciones de la Universidad de Cádiz, 2011, ISBN 978-84-9828-348-8, pp. 221-225, (2011).

[6] Janáček, J., Kvet, M.: *Public service system design with disutility relevance estimation*. In: Proceedings of the 31 st international conference "Mathematical Methods in Economics", September 11-13, 2013, Jihlava, Czech Republic, pp 332 – 337

[7] Ogryczak, W., Sliwinski, T.: *On Direct Methods for Lexicographic Min-Max Optimization*. In M. Gavrilova et al. (Eds.): ICCSA 2006, LNCS 3982, pp. 802-811, Springer-Verlag Berlin Heidelberg, 2006.

Modification of the AHP based model for evaluating artistic production of Czech colleges

Věra Jandová[1], Jan Stoklasa[2], Jana Talašová[3]

Abstract. The Registry of Artistic Performances (RUV in Czech) is the registry where the information about the creative work outcomes of Czech colleges with artistic production are stored. A model based on the Analytic Hierarchy Process was chosen for evaluating these works of art. It has been conceived also as a quality assessment tool that would be able to provide valid information for funds distribution among the Czech colleges. In fact, money from the state budget of the Czech Republic which should be allocated to the Czech universities and colleges is currently distributed based on RUV as one of the parameters.

After the pilot run of RUV, a detailed analysis of the data and evaluations stored in RUV was performed. It confirmed that the model and the structure of the categories are in accordance with the purpose of RUV. The analysis also inspired further development of the evaluation model which will be described in this paper. During the process of the model modifications, a new correct version of AHP in the ratings mode was proposed. This version is presented in this paper.

Keywords: Decision making, Saaty's AHP, RUV, Ratings mode, evaluation of categories

JEL classification: C44
AMS classification: 91B06

1 Registry of Artistic Performances

The Registry of Artistic Performances (RUV in Czech) is a registry where the information about creative work outcomes of Czech colleges with artistic production are stored. A model based on the Analytic Hierarchy Process (AHP) was chosen for evaluating these works of art. Although the model has been proposed as a quality assessment tool that would provide valid information for funds distribution, an important idea behind it is to promote excellence in artistic production. The art production is divided into several categories for the purpose of inserting the outcomes into the RUV. According to this classification, every piece of art is assigned a certain score. The scores of all art production of every college are summed and this value is considered as a measure of the quality of the production activity of the college in the field of arts. The scores given to colleges for the artistic production in the past five years stored in RUV have been used as one of the parameters for distributing money from the state budget of the Czech Republic among the Czech colleges since 2012 (see [3]).

The artistic production in the Czech Republic is divided into the following seven segments for the purpose of RUV - architecture, design, film, fine arts, literature, music and theatre. The evaluation model is uniform for all of these fields. The pieces of art are classified according to three criteria. Every criterion distinguishes three different levels which are labeled by the capital letters:

1. Relevance/significance of a new piece of art/performance: **A** - crucial significance, **B** - containing numerous important innovations, **C** - pushing forward modern trends.

[1]Palacký University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic, vera.jandova@upol.cz
[2]Palacký University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic, jan.stoklasa@upol.cz
[3]Palacký University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic, jana.talasova@upol.cz

2. Extent of a piece of art/performance: **K** - large, **L** - medium, **M** - limited.

3. Institutional or media reception/impact of a piece of art/performance: **X** - international, **Y** - national, **Z** - regional.

The value of the criterion Relevance/significance is determined by the judgments of experts. The levels of the criterion Extent are relatively clearly defined for each segment of the artistic production and the levels of the criterion Institutional reception are given by lists of institutions which are compiled specifically for each segment. The resulting category is described by a triplet of capital letters, e.g. AKX, BLY or CMZ.

2 Analytic Hierarchy Process in the Ratings Mode

The AHP model [5, 6] was selected as the evaluation method in RUV. The AHP in general is based on creating a hierarchical structure of a problem and using pairwise comparison matrices of elements. The Saaty's scale is used for setting these matrices and the consistency condition is supervised. The details how to calculate weights of criteria and evaluations of alternatives in the classical AHP are given in [5]. In this section, a special case of AHP denoted by Saaty as AHP in the Ratings mode [4] will be discussed.

2.1 Structure of the model

Let us consider a set of alternatives with characteristics that are difficult to measure or quantify (e.g. works of art) that need to be evaluated according to several criteria. Ratings of these alternatives can be given indirectly through scores of general categories to which these alternatives are assigned. The categories are created by combinations of linguistically expressed levels of criteria. Scores of these categories are computed by an AHP-based method and these scores are assigned to the respective alternatives.

Let us consider m criteria $C_1, C_2, \ldots, C_m$. The linguistically expressed levels $L_1^j, L_2^j, \ldots, L_{n_j}^j$ of a criterion C_j, $j = 1, 2, \ldots, m$, will be numbered according to their evaluation, for simplicity we will use the rank of the level to denote it. (The example of criterion levels can be *Large, Medium* and *Limited* for the criterion Extent.) A general category will be identified by an m-tuple $(i_1, i_2, \ldots, i_m)$, $i_k \in \{1, 2, \ldots, n_k\}$, $k = 1, 2, \ldots, m$. The importance of the criteria will be denoted by the normalized weights $w_1, w_2, \ldots, w_m$, $w_j \geq 0$, $j = 1, 2, \ldots, m$ and $\sum_{j=1}^m w_j = 1$.

The weights are derived through the Saaty's pairwise comparison matrix $S = \{s_{ik}\}_{i,k=1}^m$. The Eigenvector method is used for this computation in the AHP (see [5]). In a similar way the normalized scores $p_1^j, p_2^j, \ldots p_{n_j}^j$, $p_i^j \geq 0$, $i = 1, 2, \ldots, n_j$, $\sum_{i=1}^{n_j} p_i^j = 1$ for all $j = 1, 2, \ldots, m$, for the levels of criteria can be computed from the Saaty's matrix $S^j = \{s_{ik}^j\}_{i,k=1}^m$ for all $j = 1, 2, \ldots, m$. The rating of a category $(i_1, i_2, \ldots, i_m)$, $i_k \in \{1, 2, \ldots, n_k\}$, $k = 1, 2, \ldots, m$, will be denoted $R_{i_1, i_2, \ldots, i_m}$ and according to the AHP method, it will be computed as a weighted average $R_{i_1, i_2, \ldots, i_m} = \sum_{j=1}^m w_j R_{i_1, i_2, \ldots, i_m}^j$ where $R_{i_1, i_2, \ldots, i_m}^j$ is the evaluation of the category $(i_1, i_2, \ldots, i_m)$ with respect to criterion C_j computed based on the score $p_{i_j}^j$ of the i_j-th level of the criterion C_j.

2.2 Deriving the ratings of categories

In this section, two approaches that were proposed by Saaty to compute the scores of the categories with respect to the criteria will be presented. Both of them will be analyzed and their shortcomings pointed out. In the end, a proper formula for the calculation of $R_{i_1, i_2, \ldots, i_m}^j$, $j = 1, 2, \ldots, m$ will be introduced. The corresponding proper formula for ratings of categories (3) will be provided based on the formula for the calculation of $R_{i_1, i_2, \ldots, i_m}^j$.

In [5] Saaty described the AHP in the ratings mode. He set $R_{i_1, i_2, \ldots, i_m}^j = p_{i_j}^j$ and thus

$$R_{i_1, i_2, \ldots, i_m} = \sum_{j=1}^m w_j p_{i_j}^j. \tag{1}$$

The scores computed by this formula are affected by the number of the criterion levels. As a consequence, the criterion with the minimal number of levels would have greater effect on the overall scores of the categories than it should have according to the given weights. Under the same number of levels (let us consider n) for all criteria, the sum of the ratings of all categories will not be equal to 1, as it is customary for the overall evaluations in AHP (the sum is equal to n^{m-1}).

Saaty replied to this problem in [4, 6] by proposing to perform standardization of the scores of the levels of criteria - each weight of the level computed by the Eigenvector method is divided by the maximum of the computed weights. Under our notation, this can be summarized as $R^j_{i_1,i_2,\ldots,i_m} = p^j_{i_j} / \max\{p^j_i; i = 1,2,\ldots,n_j\}$ (it is $R^j_{i_1,i_2,\ldots,i_m} = p^j_{i_j}/p^j_1$ in our case) and thus

$$R_{i_1,i_2,\ldots,i_m} = \sum_{j=1}^{m} w_j \frac{p^j_{i_j}}{\max\limits_{i=1,2,\ldots,n_j} p^j_i}. \tag{2}$$

The scores of the categories calculated by the formula (2) are called absolute evaluations by Saaty [4]. Even if the maximum level score is 1 for each criterion, the minimum level score for C_j is $p^j_{n_j}/p^j_1$, that means it is possibly different for different criteria. Therefore, it is strange to talk about an absolute evaluation and it is also weird to compute a weighted average of evaluations defined on different scales. In addition, the type of the evaluation is far from the original one applied in the classical AHP.

We propose a solution that is a direct application of the AHP method on categories defined by combinations of criteria levels. Strictly speaking this would mean that Saaty's matrices should be constructed for the comparison of all the categories under each criterion (the dimension of such matrices would be $n_1 \cdot n_2 \cdots \cdot n_m$). It is however not necessary to create such a big matrices and compute the scores of categories with respect to criteria, $R^j_{i_1,i_2,\ldots,i_m}$, $j = 1,2,\ldots,m$. We can use the scores of the criteria levels $p^j_1, p^j_2, \ldots p^j_{n_j}$, $j = 1,2,\ldots,m$. Thereafter during the process of normalization, we need to consider the fact that each of the scores of the levels of the criterion C_j appears in $\prod_{k=1,k\neq j}^{m} n_k$ categories for each $j = 1,2,\ldots,m$. Therefore, the scores of the categories with respect to the criteria are equal to $R^j_{i_1,i_2,\ldots,i_m} = p^j_{i_j} / \prod_{k=1,k\neq j}^{m} n_k$ and

$$R_{i_1,i_2,\ldots,i_m} = \sum_{j=1}^{m} w_j \frac{p^j_{i_j}}{\prod\limits_{k=1,k\neq j}^{m} n_k}. \tag{3}$$

Notice that in this case $\sum_{i_1,\ldots,i_m} R_{i_1,i_2,\ldots,i_m} = 1$. This alternative of the AHP in Ratings mode is clearly a special case of the AHP for criteria whose values are expressed linguistically without an explicit mathematical interpretation of these linguistic values. This version of the AHP method can also be used in the case when we combine quantitative and qualitative criteria and the evaluations according to the qualitative criteria are provided in this manner. When normalizing the evaluations under such qualitative criteria, it is necessary to take into account how many times the given level is presented.

3 Evaluation model for the RUV

3.1 The original model

The AHP was used to determine the scores of the RUV categories. Due to the existence of interactions among the three criteria (Relevance/significance, Extent and Media/institutional reception), scores were assigned directly to the triplets of the criteria levels which represent 27 categories of the artistic production. The second reason for creating directly the matrix of the order 27 was the effort to obtain a high-quality information about the priorities between the given categories from the experts from the fields of art. Examples of real-life works of art were assigned to every category in all segments. With regard to this, experts compared these representatives instead of the abstract categories. The computation of scores of the 27 categories is described in detail in [7, 8]. First, the ordering of the categories was given by the Pairwise comparison method. Next, the scores were determined from the Saaty's matrix using the Eigenvector method. The results were subsequently rescaled to set the value for the best category AKX to 305 (to achieve comparability with the evaluation of R&D outputs in the Czech Republic [3]).

The resulting scores of the categories are summarized in Table 1. The consistency ratio of the given Saaty's matrix CR was 0.1996 (the intensities of preferences were given by a consensus of expert opinions, the matrix was of a large order). The Saaty's matrix however fulfills the weak consistency condition introduced in [8].

#	CAT.	SCORE
1	AKX	305
2	AKY	259
3	AKZ	210
4	ALX	191
5	AMX	174

#	CAT.	SCORE
6	ALY	138
7	ALZ	127
8	BKX	117
9	AMY	97
10	AMZ	90

#	CAT.	SCORE
11	BKY	79
12	BKZ	66
13	BLX	62
14	BMX	48
15	BLY	44

#	CAT.	SCORE
16	BLZ	40
17	BMY	37
18	BMZ	31
19	CKX	26
20	CLX	24

#	CAT.	SCORE
21	CKY	19
22	CKZ	17
23	CMX	16
24	CLY	12
25	CLZ	10

#	CAT.	SCORE
26	CMY	9
27	CMZ	8

Table 1 Scores of the categories in the original model

3.2 Further development of the model

In the year 2013, the software application RUV was used for the first time to collect data on the artistic production in the Czech Republic. 3 902 works of art were gathered in the database and were evaluated on the base of the method introduced in the previous subsection. After this pilot run, a detailed analysis of the artistic results and their evaluations stored in RUV was performed. It confirmed that the model and the structure of the categories are in accordance with the purpose of RUV. On the other hand, on the basis of data collected in RUV, several possible adjustments of the methodology were proposed. At present time, these suggestions are being analyzed.

One of the suggestions included adding a new level of criterion Extent: J - *collective artistic production of a very large extent*. Another proposal was concerned with the reduction of the scores of categories of the A.Z type. These categories are strange as a low institutional or media reception (Z) of the piece of art that is evaluated by a group of expert as highly relevant or significant (A) is very improbable (yet such works of art can be found in RUV). If a piece of art is really of crucial significance, its reception can be expected to rise in the next period (achieving at least Y after the re-evaluation according to the rules of RUV). For the same reason, it would be appropriate to slightly reduce the scores of categories B.Z as well. In this context it is also reasonable to consider how high the scores of the categories C.X should be.

The requirement to include a new level of the Extent criterion (J) into the model means to add 9 more categories: AJX, AJY, ..., CJZ. The matrix 27x27 would expand to 36x36 and the number of needed pairwise comparisons would rise from 351 to 630. Such a large matrix is not convenient for the experts and we cannot expect its consistency to be high. This issue might be solved by dividing the model into two smaller ones - the interactions of Extent with the other two criteria are significantly weaker than the interactions between the criteria Relevance and Reception. Even when the model is split into smaller ones, the comparison of abstract categories is not a trivial task for the experts from the field of arts. The mathematical team of the RUV project has therefore decided to utilize the basic information on priorities from the original 27x27 matrix and to propose a solution based on this information to maintain maximum consistency. Two matrices for the determination of scores were constructed. The first one for Extent (with four levels: J, K, L and M) and the second one for the two interacting criteria Relevance and Reception (with nine levels: AX, AY, ..., CZ). A third Saaty's matrix was defined for the weights of the two groups of criteria.

Table 2 presents the matrix for weights determination and the matrix for computation of the levels scores of Extent. In the second matrix, the preference intensities were allocated evenly along the Saaty's multiplicative scale and a very good consistency was achieved ($CR = 0.0426$).

comparison of criteria	RELEVANCE & RECEPTION	EXTENT	WEIGHTS
RELEVANCE & RECEPTION	1	3	0.75
EXTENT		1	0.25

	J	K	L	M	SCORES	CAT.	SCORES
J	1	3	5	7	0.5650	.J.	0.0628
K		1	3	5	0.2622	.K.	0.0291
L			1	3	0.1175	.L.	0.0131
M				1	0.0553	.M.	0.0061

Table 2 Comparison of the criteria and comparison of the levels of the criterion Extent

The matrix for comparing pairs of criteria levels for the group of criteria Relevance and Institutional reception was constructed in two steps. At first, the matrix was generated based on the following simple rule:

Each shift in Relevance by one level means a shift on the Saaty's scale by two levels in the respective direction and each shift in Reception by one level means a shift on the Saaty's scale by one level in the respective direction. Shifts are considered on the Saaty's scale $1/9, 1/7, 1/5, 1/3, 1, 3, 5, 7, 9$ against indifference, i.e. preference intensity 1. Naturally the resulting preference intensity can be maximally 9 and minimally 1/9. As an example, let us consider we want to determine the preference intensity between AY and BX. Then the shift from A to B represents +2 levels and shift from Y to X represents -1 level. The preference intensity is therefore $2-1=1$ level above indifference, which means the resulting preference intensity is between AY and BX is 3.

The matrix constructed according to this rule is presented in Table 3 on the left (CR=0.0749).

Matrix created via rule

	AX	AY	BX	AZ	BY	BZ	CX	CY	CZ	SCORES	CAT.	SCORES
AX	1	3	5	5	7	9	9	9	9	0.3503	A.X	0.0876
AY		1	3	3	5	7	7	9	9	0.2180	A.Y	0.0545
BX			1	1	3	5	5	7	9	0.1460	B.X	0.0365
AZ				1	3	5	5	7	9	0.1079	A.Z	0.0270
BY					1	3	3	5	7	0.0666	B.Y	0.0166
BZ						1	1	3	5	0.0439	B.Z	0.0110
CX							1	3	5	0.0322	C.X	0.0081
CY								1	3	0.0207	C.Y	0.0052
CZ									1	0.0144	C.Z	0.0036

Final matrix

	AX	AY	BX	AZ	BY	BZ	CX	CY	CZ	SCORES	CAT.	SCORES
AX	1	3	5	7	7	9	9	9	9	0.3759	A.X	0.0940
AY		1	3	5	5	8	8	9	9	0.2372	A.Y	0.0593
BX			1	3	3	6	6	7	9	0.1405	B.X	0.0351
AZ				1	1	4	4	5	7	0.0756	A.Z	0.0189
BY					1	4	4	5	7	0.0756	B.Y	0.0189
BZ						1	1	2	4	0.0297	B.Z	0.0074
CX							1	2	4	0.0297	C.X	0.0074
CY								1	3	0.0216	C.Y	0.0054
CZ									1	0.0144	C.Z	0.0036

Table 3 Pairwise comparison matrices for the group of criteria Relevance and Reception

In the next step, the interactions between the criteria Relevance and Reception are reflected. The preferences of the strange combinations of the criteria levels AZ, BZ and CX are lowered. In the case of AZ, this is done by one level on the previously used Saaty's scale and in the case of BZ and CX by half a level, see Table 3 for the resulting matrix - on the right (CR=0.0740).

Based on the Saaty's matrices presented in Table 2 and the final matrix in Table 3 the criteria-group weights and the scores of the levels of the groups of criteria were computed using the Eigenvector method. The scores of the 36 categories were computed by the formula (3) and transformed to again 305 for the best category. The results are presented in Table 4.

On the presented application, we have demonstrated how to improve the consistency of the solutions based on the pairwise comparison matrices of large order and how to reflect additional requests of the evaluators. It is interesting to see that the modification of the model which consists of adding a new level of the criterion Extent and dividing the model into two smaller ones, caused the ratio of the scores of the best and the worst category to get smaller. Testing computations show a systematic tendency of this difference to get smaller when a large problem is split into several smaller ones. In the case of such evaluation models, where excellence should be promoted, such a decrease of the difference between the ratings of the best and the worst category might not be desirable. If adding of a new level of criterion Extent would be necessary in RUV, the matrix 36x36 would have to be constructed. The procedure of filling the preference intensities into the Saaty's matrix would be as following: A majority of the elements of the matrix could be generated using similar simple rules as were given above. The weak consistency condition would be used for completing the rest of the matrix. Nevertheless, it is also possible to avoid including the J level by re-definition of the existing levels of Extent which would not raise the number of comparisons.

4 Conclusion

The model for evaluating creative work outcomes of Czech art colleges and other faculties with creative artistic production was summarized. State-of-the-art of the development of the AHP based model for the Registry of Artistic Performances was presented. After a detailed analysis of the artistic results and their evaluations stored in RUV, several possible adjustments of the methodology were proposed. The aim of the mathematical team was to find a model which would consider these suggestions, which would allow

#	CAT.	SCORE
1	AJX	305
2	AKX	275
3	ALX	261
4	AMX	255
5	AJY	213
6	AKY	183

#	CAT.	SCORE
7	ALY	169
8	AMY	163
9	BJX	149
10	BKX	119
11	AJZ	106
12	BJY	106

#	CAT.	SCORE
13	BLX	105
14	BMX	99
15	AKZ	76
16	BKY	76
17	BJZ	75
18	CJX	75

#	CAT.	SCORE
19	CJY	70
20	CJZ	65
21	ALZ	62
22	BLY	62
23	AMZ	56
24	BMY	56

#	CAT.	SCORE
25	BKZ	45
26	CKX	45
27	CKY	40
28	CKZ	35
29	BLZ	31
30	CLX	31

#	CAT.	SCORE
31	CLY	26
32	BMZ	25
33	CMX	25
34	CLZ	21
35	CMY	20
36	CMZ	15

Table 4 Scores of the categories in the adjusted model

to preserve the preference intensities given by the experts, which would allow to fill in the rest of the preference intensities in a simple way and which would lead to improving the value of the Consistency ratio for the created Saaty's matrix. The AHP in Ratings mode was considered in the modified model for evaluating artistic production gathered in RUV which was exhibited in this article.

From the general mathematical point of view this paper introduces a proper way of determining the scores of categories in the AHP in Ratings mode. The proposed formula (3) eliminates several drawbacks of the computations of these scores published previously by Saaty.

Acknowledgements

Supported by the grant IGA_PrF_2014028 *Mathematical models* of the Internal Grant Agency of Palacky University in Olomouc and by the grant GA 14-02424S *Methods of operations research for decision support under uncertainty* of the Grant Agency of the Czech Republic.

References

[1] Alonso, J. A. and Lamata, M. T.: Consistency in the Analytic Hierarchy Process: A New Approach, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **14** (2006), 445–459.

[2] Jandová, V. and Talašová, J.: Weak consistency: A New Approach to Consistency in the Saaty's Analytic Hierarchy Process. *Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica* **52** (2013), 71–83.

[3] Ministerstvo školství, mládeže a tělovýchovy: *Zásady a pravidla financování veřejných vysokých škol pro rok 2012*. Praha, Česká republika, 2011.

[4] Saaty, T. L.: Decision making with the analytic hierarchy process, *International Journal of Services Sciences* **1** (2008), 83–98.

[5] Saaty, T. L.: How to make a decision: The Analytic Hierarchy Process, *European Journal of Operational Research* **48** (1990), 9–26.

[6] Saaty, T. L.: *Fundamentals of Decision Making and Priority Theory With the Analytic Hierarchy Process*. RWS Publications, Pittsburgh, 2000.

[7] Stoklasa, J., Jandová, V. and Talašová, J.: Weak consistency in Saaty's AHP - evaluating creative work outcomes of Czech Art Colleges. *Neural Network World* **23** (2013), 61–77.

[8] Talašová, J. and Stoklasa, J.: A Model for Evaluating Creative Work Outcomes at Czech Art Colleges. In *Proceedings of the 29th International Conference on Mathematical Methods in Economics - Part II*, (2011), 698-703.

Robust regularized cluster analysis for high-dimensional data

Jan Kalina[1], Katarína Vlčková[2]

Abstract. This paper presents new approaches to the hierarchical agglomerative cluster analysis for high-dimensional data. First, we propose a regularized version of the hierarchical cluster analysis for categorical data with a large number of categories. It exploits a regularized version of various test statistics of homogeneity in contingency tables as the measure of distance between two clusters.

Further, our aim is cluster analysis of continuous data with a large number of variables. Various regularization techniques tailor-made for high-dimensional data have been proposed, which have however turned out to suffer from a high sensitivity to the presence of outlying measurements in the data. As a robust solution, we recommend to combine two newly proposed methods, namely a regularized version of robust principal component analysis and a regularized Mahalanobis distance, which is based on an asymptotically optimal regularization of the covariance matrix. We bring arguments in favor of the newly proposed methods.

Keywords: cluster analysis, robust data mining, big data, regularization.

JEL classification: C14
AMS classification: 62G35

1 Cluster analysis

Cluster analysis represents a general information extraction methodology allowing to reveal the multivariate structure of given data and to divide multivariate data to subpopulations. It is often used as an exploratory technique for complex multivariate data and can be also interpreted as a dimensionality reduction technique. It covers a wide range of flexible methods allowing the user to select one of available algorithms with numerous parameters to adjust the whole computation to the data. Clearly, the results strongly depend on the choice of the particular method.

Hierarchical clustering represents one of the most commonly used methods of unsupervised learning [5, 12]. It is an agglomerative (bottom-up) method starting with individual objects as individual clusters and merges recursively a selected pair of clusters into a single cluster. The method repeatedly searches for the pair of clusters with the largest similarity, until a given stopping rule is fulfilled. Numerous choices of distances are available to measure the similarity between two clusters, while some of them may be rather interpreted as dissimilarity. Hierarchical clustering has become a popular procedure in economic applications, e.g. in strategic management or credit risk management. While cluster analysis within statistics is used mainly for prediction, its main aim in economic applications is shifted towards variable selection and clear interpretation of multivariate data.

Standard hierarchical cluster analysis is known to suffer from the presence of outliers in the data. Therefore, we propose new approaches to robust cluster analysis for high-dimensional data contaminated by outlying measurements. Because cluster analysis is always based on some measure of distance between two clusters, we propose new measures of distance between two clusters and one of standard algorithms may be used to perform the cluster analysis itself.

The common idea of cluster analysis methods for high-dimensional data is regularization, which can be described as exploiting additional information, assumptions, or penalization to solve ill-posed or insol-

[1]Institute of Computer Science AS CR, Pod Vodárenskou věží 2, 182 07 Praha 8, Czech Republic, kalina@cs.cas.cz
[2]Institute of Computer Science AS CR, Pod Vodárenskou věží 2, 182 07 Praha 8, Czech Republic, vlckova@cs.cas.cz

uble high-dimensional problems [5]. It ensures a numerical feasibility and stability of the computation. Regularized estimates are biased, but their mean square error is drastically reduced compared to that of classical unbiased estimates. Yao *et al.* [19] proposed a k-means clustering method based on a shrinkage version of the correlation coefficient. Gao and Hitchcock [3] proposed a k-means cluster analysis based on the Euclidean distance of each observation from a shrinkage mean of a particular cluster, while assuming the covariance matrix in each cluster to be regular. This is inappropriate and a numerically stable solution could be obtained with a regularized covariance matrix (e.g. [14]), which commonly has the form of a linear combination of the empirical covariance matrix $\mathbf{S}$ and a regular (e.g. diagonal) target matrix.

Cluster analysis as a method for finding clusters of variables in a multivariate setting is particularly important for a high-dimensional context, when the number of variables p largely exceeds the number of observations n (i.e. $n \ll p$). Unfortunately, traditional approaches to cluster analysis are unsuitable for this context and attention has been paid to specific hierarchical cluster analysis for high dimensional data only recently [12], primarily in bioinformatics [10], although they represent an important problem also in the analysis of economic data [11]. While the methods are reported to robust from the point of view of numerical mathematics (i.e. computationally stable), regularization does not bring statistical robustness with respect to the presence of outliers. Thus, robustness to outliers remains to be an important requirement expected from reliable data mining methods for high-dimensional data.

Section 2 of this paper proposes a hierarchical agglomerative cluster analysis method for multinomial data with a large number of categories. Section 3 proposes a hierarhical agglomerative cluster analysis method for continuous data, which combines a highly robust dimensionality reduction and a regularization of the covariance matrix. Finally, Section 4 concludes the paper.

2 Hierarchical agglomerative cluster analysis for multinomial data

The Pearson χ^2 statistic and the likelihood ratio statistic G^2 (i.e. the deviance) are commonly used in hierarchical agglomerative cluster analysis method for multinomial data. We discuss their disadvantages for data with a large number of categories. We propose a new method of hierarchical agglomerative cluster analysis for multinomial data. It uses a shrinkage regularization to derive an alternative version of the common test statistics. For this purpose, we derive the optimal value of the regularization parameter.

If one or more cells in the contingency table is equal to zero, the likelihood ratio test statistic G^2 cannot be computed. The χ^2 statistic can be computed, but still a zero count is heavily influential on the value of χ^2. Other disadvantages of the presence of structural or sampling zero counts are well known e.g. in the context of generalized linear models [1], where G^2 is more commonly applied than the χ^2. Therefore, Agresti [1] recommends to always replace zero counts by small constants (e.g. 1/2) in the computation of χ^2. More sophisticated regularization techniques have been proposed to modify the χ^2 or G^2 statistics for data with zero counts, which ensure robustness to influential data. While robust methods for the analysis of categorical data have been studied rather rarely [9], Buonaccorsi [2] warned that outliers appear in discrete data quite commonly, namely in the form of measurement errors.

Now, we will propose a regularized version of the χ^2 and G^2 statistics with an analytical expression for asymptotically optimal regularization parameter. At the same time, the regularized χ^2 statistic can be interpreted as a sum of standardized residuals computed for the regularized test statistic. For discrete data, regularization is known to reduce the influence of outlying measurements and to ensure robustness [13, 6].

Let us observe a random sample of n random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ following a multinomial distribution with J categories, where the value of J is large. We use the notation $\mathbf{X}_l = (\mathbf{X}_{l1}, \ldots, \mathbf{X}_{lJ})^T$ for $l = 1, \ldots, n$. Cluster analysis applied on the data requires to measure the distance between observations $\mathbf{X}_l$ and $\mathbf{X}_{l'}$ for $l = 1, \ldots, n$, $l' = 1, \ldots, n$ and $l \neq l'$. Such tests are commonly based on the χ^2 or G^2 test statistic computed from the contingency table

	Group 1	Group 2	$\cdots$	Group J
Observation l	X_{l1}	X_{l2}	$\cdots$	X_{lp}
Observation l'	$X_{l'1}$	$X_{l'1}$	$\cdots$	$X_{l'p}$

. (1)

Our aim is to obtain a regularized version of the Pearson χ^2 statistic using the scheme

$$\chi^{*2} = \sum \frac{(O - E^*)^2}{E^*}, \tag{2}$$

where the sum goes over all counts in (1) and compares each of the observed counts O with a regularized version of the corresponding expected count E^*. In an analogous way, a regularized version of the likelihood ratio test statistic will be obtained in a schematic notation

$$G^{*2} = \sum O \log \frac{O}{E^*}. \tag{3}$$

We use the notation $X_{\cdot j} = X_{1j} + X_{2j}$ for $j = 1, \ldots, J$ and $X_{\cdot\cdot} = \sum_{j=1}^{J}(X_{1j} + X_{2j})$. We propose to estimate the parameters of the multinomial distribution by regularized (biased) estimators in the form

$$p_j^* = (1 - \lambda^*)\frac{X_{1j}}{X_{\cdot j}} + \lambda^* \frac{X_{1\cdot}}{X_{\cdot\cdot}}, \quad j = 1, \ldots, J, \quad \lambda \in [0, 1]. \tag{4}$$

This can be interpreted as a regularized version of the maximum likelihood estimator, which is shrunken to the overall maximum likelihood estimator across categories. Replacing maximum likelihood estimates by (4), we can express the χ^{*2} statistic (2) as

$$\chi^{*2} = \sum_{j=1}^{J} \left[\frac{(X_{1j} - X_{\cdot j} p_j^*)^2}{X_{\cdot j} p_j^*} + \frac{(X_{2j} - X_{\cdot j}(1 - p_j^*))^2}{X_{\cdot j}(1 - p_j^*)} \right]. \tag{5}$$

The regularized likelihood ratio test statistic (3) can be expressed as

$$G^{*2} = \sum_{j=1}^{J} \left[X_{1j} \log \frac{X_{1j}}{X_{\cdot j} p_j^*} + X_{2k} \log \frac{X_{2j}}{X_{\cdot j}(1 - p_j^*)} \right]. \tag{6}$$

Besides, we can derive an asymptotically optimal value of λ, which minimizes the mean square error among all values of (4) over $\lambda \in (0, 1)$. following the asymptotic reasoning of [11] for $p \to \infty$. This optimal value of λ will be denoted as λ^* and can be expressed as

$$\min\{\lambda^*, 1\} \tag{7}$$

where

$$\lambda^* = \frac{1 - \sum_{j=1}^{J} \left(\frac{X_{1j}}{X_{\cdot j}}\right)^2}{(n-1)\sum_{j=1}^{J} \left(\frac{X_{1\cdot}}{X_{\cdot\cdot}} - \frac{X_{1j}}{X_{\cdot j}}\right)^2}. \tag{8}$$

We may interpret χ^{*2} not only as a test statistic, but at the same as a basis for obtaining a regularized version of the χ^2 distance. The phi coefficient φ is a commonly based measure in the context of cluster analysis for discrete data and we propose to measure the distance between observations $\mathbf{X}_l$ and $\mathbf{X}_{l'}$ by a regularized version of φ in the form

$$\varphi^* = \sqrt{\frac{\chi^{*2}}{n}}. \tag{9}$$

The cluster analysis can be performed by one of existing algorithms, where $\mathbf{X}_l$ and $\mathbf{X}_{l'}$ are representants of two clusters chosen according to a selected linkage criterion. Thus, φ^* is not only the distance between two observations, but plays the role of a distance between two clusters.

3 Hierarchical agglomerative cluster analysis for continuous data

In this section, we propose new methods for hierarchical agglomerative cluster analysis for continuous high-dimensional data using a suitable regularization. After a robust dimensionality reduction, we propose a robust cluster analysis procedure based on a robust distance between two clusters. We discuss the choice of a suitable robust measure for this purpose and we propose a regularized Mahalanobis distance. The resulting two-stage cluster analysis method combines regularization and robustness.

Our first proposal is a robust dimensionality reduction method. The principal component analysis (PCA) is known to suffer from the presence of outliers. Besides, the PCA does not need a regularization to improve numerical stability, but requires a regularization in order to obtain a sparse approximation of the data by a low-rank matrix [16]. Therefore, we propose a new approach to the PCA in the form of combining regularization with robustness. The idea of implicit weighting of individual observations was originated in the idea of the least weighted squares (LWS) regression [18, 8]. Now, we propose a robust dimensionality reduction method suitable for high-dimensional data.

Algorithm 1. (Robust regularized principal component analysis.)

Let us consider p-dimensional data vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

1. Compute the LWS-correlation matrix

$$\mathbf{R}_{LWS} = \left(R_{ij}^{LWS}\right)_{i,j=1}^{p}, \tag{10}$$

   where R_{ij}^{LWS} is equal to the LWS-correlation coefficient between data vectors $(X_{1i}, \ldots, X_{ni})^T$ and $(X_{1j}, \ldots, X_{nj})^T$.

2. Compute the LWS-covariance matrix

$$\mathbf{S}_{LWS} = \left(S_{ij}^{LWS}\right)_{i,j=1}^{p} \tag{11}$$

   as

$$S_{ij}^{LWS} = R_{ij}^{LWS}\sqrt{S_{ii}^{LWS}S_{jj}^{LWS}}, \tag{12}$$

   where S_{ii}^{LWS} and S_{jj}^{LWS} are robust estimators of variance based on the LWS [7].

3. Compute the value of the regularization parameter $\lambda^* \in [0, 1]$ as

$$\lambda^* = \frac{2\sum_{i=2}^{p}\sum_{j=1}^{i-1}\widehat{\mathsf{var}}(S_{ij}^{LWS})}{2\sum_{i=2}^{p}\sum_{j=1}^{i-1}(S_{ij}^{LWS})^2}, \tag{13}$$

   where $\widehat{\mathsf{var}}(S_{ij}^{LWS})$ denotes the empirical variance of values all S_{ij}^{LWS} for $i = 1, \ldots, p$ and $j = 1, \ldots, p$.

4. Compute the regularized robust covariance matrix

$$\mathbf{S}^* = (1 - \lambda^*)\mathbf{S}_{LWS} + \lambda^*\boldsymbol{\mathcal{I}}. \tag{14}$$

5. Compute the regularized robust correlation matrix

$$\mathbf{R}^* = \left(R_{ij}^*\right)_{i,j=1}^{p}, \tag{15}$$

   where

$$R_{ij}^* = \frac{S_{ij}^*}{\sqrt{S_{ii}^*S_{jj}^*}}, \quad i, j = 1, \ldots, p. \tag{16}$$

6. Compute the eigenvalues of $\mathbf{R}^*$ and denote them as $\theta_1^*, \ldots, \theta_p^*$. The corresponding eigenvectors will be denoted as $\mathbf{z}_1^*, \ldots, \mathbf{z}_p^*$.

7. The observations $\mathbf{X}_i$ are replaced by the set of the first r principal components $\mathbf{z}_1^{*T}\mathbf{X}_i, \ldots, \mathbf{z}_r^{*T}\mathbf{X}_i$ for $i = 1, \ldots, n$, where a given rule for the selection of r is used, e.g. to find r as the minimal integer fulfilling $\sum_{j=1}^{r}\theta_j^* \geq 0.9$ and $r \leq p$.

The highly robust dimensionality reduction of Algorithm 1 is based on (14), which is a robust analogy of various regularization techniques for the empirical covariance matrix of [14]. (13) represents a robust analogy of the approach of [15], who derived an optimal regularization parameter minimizing the mean square error of $\mathbf{S}^*$ over $\lambda \in [0, 1]$.

Consequently, one of standard algorithms for cluster analysis may be used in a classical way. Because regularization does not ensure robustness properties for continuous data, it is desirable to consider a robust approach to the consequent cluster analysis [10]. If the user wishes to use the Pearson correlation

coefficient as the measure of distance between two clusters, then its robust alternative should be a preferable choice. We proposed a robust correlation coefficient r_{LWS} based on the idea of implicit weighting of individual observations in [7].

Further, we define the regularized Mahalanobis distance, applicable to hierarchical cluster analysis. Let us assume two groups of observations $\mathbf{X}_1 = (X_{11}, \ldots, X_{1n_1})^T$ and $\mathbf{X}_2 = (X_{21}, \ldots, X_{2n_2})^T$ as the distance $d(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2)$, where $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$ are the means of both groups. Let $\mathbf{S}$ denote the pooled covariance matrix computed across both groups.

Algorithm 2. (Regularized Mahalanobis distance.)

1. Compute the eigendecomposition of $\mathbf{S}$ as
$$\mathbf{S} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T, \tag{17}$$
where
$$\mathbf{D} = diag\{\theta_1, \ldots, \theta_p\} \quad \text{and} \quad \mathbf{Q}^T = \mathbf{Q}^{-1}. \tag{18}$$
2. For
$$\hat{\lambda} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} \widehat{\text{var}}(S_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{p}(S_{ij} - \ _{[i=j]})^2} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} \widehat{\text{var}}(S_{ij})}{2\sum_{i=2}^{p}\sum_{j=1}^{i-1} S_{ij}^2 + \sum_{i=1}^{p}(S_{ii}-1)^2}, \tag{19}$$
compute
$$\mathbf{D}^* = \mathsf{diag}\{(1-\hat{\lambda})\theta_1 + \hat{\lambda}, \ldots, (1-\hat{\lambda})\theta_p + \hat{\lambda}\}. \tag{20}$$
3. Compute
$$(\mathbf{D}^*)^{-1} = \mathsf{diag}\left\{\left((1-\hat{\lambda})\theta_1 + \hat{\lambda}\right)^{-1}, \ldots, \left((1-\hat{\lambda})\theta_p + \hat{\lambda}\right)^{-1}\right\}. \tag{21}$$
4. Compute
$$d(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2) = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^T \mathbf{Q}\,(\mathbf{D}^*)^{-1}\,\mathbf{Q}^T(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2). \tag{22}$$

4 Conclusions

This paper proposes new methods for hierarchical cluster analysis for high-dimensional discrete as well as continuous data.

In Section 2, a regularized method for multinomial data is proposed. To the best of our knowledge, the proposal represents a first attempt to formulate a regularized version of the popular χ^2 test statistic, exploiting an explicit expression for the asymptotically optimal value of the regularization parameter. The method is tailor-made for a large number of categories J without supposing an effect for a small or moderate J. The method contains an implicit vaguely formulated assumption that $X_{1\cdot}/X_{\cdot\cdot}$ is a relatively suitable estimate of the probability of success in each category. Thus, the performance cannot be good if these probabilities are very different from each other. Our belief that χ^{*2} and G^{*2} do not suffer from the presence of influential data values is supported by the experience of [6], who applied regularization to estimate entropy from multinomial data with a large number of categories.

In Section 3, a regularization method for continuous data is proposed. The proposed approach can be described as a two-stage robust cluster analysis method for high-dimensional data, which represents one of few available robust regularized approaches. The robustness of the method is ensured by the robustness of each step of the computation. The first step is a robust regularized dimensionality reduction and the second can be described as a robust cluster analysis, which may proceed by replacing the standard Mahalanobis distance by its regularized version in one of existing cluster analysis algorithms.

The approach of Section 3 is heuristical but based again on an asymptotically optimal value of the regularization parameter and thus represents one of first attempts to combine regularization with robustness with respect to outliers. Algorithm 2 is based on an explicit expression for eigenvalues of a regularized covariance matrix. To be specific, (20) shows the effect of the selected regularization on the eigenvalues and opens a door for alternative regularization techniques to be applied directly on the eigenvalues, extending the ideas of [14]. The idea of a regularized Mahalanobis distance, although not called by this name, has been successfully introduced to the concept of linear discriminant analysis in bioinformatics [4]. However, its performance cannot be reliable if individual variables have substantially different variances. Still it remains an open problem to investigate the relationship between regularization and statistical robustness for continuous data.

Acknowledgements

The work was supported by by the Czech Science Foundation project No. 13-17187S. The work of J. Kalina was supported also by the Czech Science Foundation project No. 13-01930S.

References

[1] Agresti, A.: *Categorical data analysis.* 2nd edn. Wiley, New York, 2002.

[2] Buonaccorsi, J. P.: *Measurement error: models, methods, and applications.* Chapman & Hall/CRC, Boca Raton, 2010.

[3] Gao, J., and Hitchcock, D.B.: James-Stein shrinkage to improve k-means cluster analysis, *Computational Statistics & Data Analysis* **54** (2010), 2113–2127.

[4] Guo, Y., Hastie, T., and Tibshirani, R.: Regularized discriminant analysis and its application in microarrays, *Biostatistics* **8** (2007), 86–100.

[5] Hastie, T., Tibshirani, R., and Friedman, J.: *The elements of statistical learning.* 2nd edn. Springer, New York, 2008.

[6] Hausser, J., and Strimmer, K.: Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks, *Journal of Machine Learning Research* **10** (2009), 1469–1484.

[7] Kalina, J.: Implicitly weighted methods in robust image analysis, *Journal of Mathematical Imaging and Vision* **44** (2012), 449–462.

[8] Kalina, J.: On multivariate methods in robust econometrics, *Prague Economic Papers* **21** (2012), 69–82.

[9] Kalina, J.: Nonlinear trend modeling in the analysis of categorical data, *The 6th International Days of Statistics and Economics Conference Proceedings (MSED 2012)*, Melandrium, Praha, pp. 1–10, ISBN 978-80-86175-86-7.

[10] Kalina, J.: Classification analysis methods for high-dimensional genetic data, *Biocybernetics and Biomedical Engineering* **34** (2014), 10–18.

[11] Ledoit, O., and Wolf, M.: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* **10** (2003), 603–621.

[12] Martinez, W.L., Martinez, A.R., and Solka, J.L.: *Exploratory data analysis with MATLAB.* 2nd edn. Chapman & Hall/CRC, London, 2011.

[13] Opgen-Rhein, R., and Strimmer, K.: Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach, *Statistical Applications in Genetics and Molecular Biology* **6** (2007), Article 9.

[14] Pourahmadi, M.: *High-dimensional covariance estimation.* Wiley, New York, 2013.

[15] Schäfer, J., and Strimmer, K.: A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, *Statistical Applications in Genetics and Molecular Biology* **4** (2005), Article 32, 1–30.

[16] Shen, H., and Huang, J. Z.: Sparse principal component analysis via regularized low rank matrix approximation, *Journal of Multivariate Analysis* **99** (2008), 1015–1034.

[17] Tibshirani, R., Walther, G., and Hastie, T.: Estimating the number of clusters in a data set via the gap statistic, *Journal of the Royal Statistical Society Series B* **63** (2001), 411–423.

[18] Víšek, J. Á.: Regression with high breakdown point. In: *Proceedings of ROBUST 2000, Summer School of JČMF,* (J. Antoch, G. Dohnal, eds.), JČMF and Czech Statistical Society, Prague, 2001, 324–356.

[19] Yao, J., Chang, C., Salmi, M. L., Hung, Y. S., Loraine, A., and Roux, S. J.: Genome-scale cluster analysis of replicated microarrays using shrinkage correlation coefficient, *BMC Bioinformatics* **9** (2008), Article 288, 1–16.

Multiobjective Stochastic Optimization Problems with Probability Constraints

Vlasta Kaňková[1]

Abstract. Multiobjective optimization problems depending on a probability measure often correspond to situations in which an economic or financial process is simultaneously influenced by a random factor and a decision parameter. Moreover it is reasonable to evaluate the process by a few objective functions and it seems reasonable to determine the "decision" with respect to the mathematical expectation of objectives. Complete knowledge of the probability measure is a necessary condition to analyze the problem. However, in applications mostly the problem has to be solved on the data base. A relationship between "characteristics" obtained on the base of complete knowledge of the probability measure and those obtained on the above mentioned data base has been already investigated in the case when constraints set is not depending on the probability measure ([9], [10]). The aim of the work will be to try to relax this condition.

Keywords: Stochastic multiobjective optimization problems, (properly) efficient solution, (strongly) convex functions, empirical estimates, Lipschitz property, constraints depending on the probability measure.

JEL classification: C44
AMS classification: 90C15

1 Introduction

To introduce a "rather general" multiobjective stochastic programming problem, let $(\Omega, \mathcal{S}, P)$ be a probability space; $\xi := \xi(\omega) = (\xi_1(\omega), \ldots, \xi_s(\omega))$ s–dimensional random vector defined on $(\Omega, \mathcal{S}, P)$; $F(:= F(z),\ z \in R^s)$, P_F and Z_F denote the distribution function, the probability measure and the support corresponding to ξ, respectively. Let, moreover, $g_i := g_i(x,\ z)$, $i = 1,\ \ldots,\ l$, $l \geq 1$ be real–valued (say, continuous) functions defined on $R^n \times R^s$; $X_F \subset X \subset R^n$ be a nonempty set generally depending on F, and $X \subset R^n$ be a nonempty "deterministic" convex set. If the symbol E_F denotes the operator of mathematical expectation corresponding to F and if for every $x \in X$ there exist finite $\mathsf{E}_F g_i(x, \xi)$, $i = 1,\ \ldots,\ l$, then a rather general "multiobjective" one–stage stochastic programming problem can be introduced in the form:

$$\text{Find} \quad \min \mathsf{E}_F g_i(x, \xi),\ i = 1,\ \ldots,\ l \quad \text{subject to} \quad x \in X_F. \tag{1}$$

The problem (1) depends on the measure P_F that usually (in applications) has to be estimated on the data base. Evidently, then an analysis have to be done with respect to an empirical problem:

$$\text{Find} \quad \min \mathsf{E}_{F^N} g_i(x, \xi),\ i = 1,\ \ldots,\ l \quad \text{subject to} \quad x \in X_{F^N}, \tag{2}$$

where F^N denotes an empirical distribution function determined by a random sample $\{\xi^i\}_{i=1}^N$ corresponding to the distribution F.

Of course by this approach we can obtain only estimates of the corresponding "theoretical" characteristics. A relationship between "theoretical" characteristics and those obtained on the data base has been already investigated in the case when the constraint set does not depend on the probability measure (see e.g. [9] and [10]). In this work we consider the case when there exist real valued convex functions

[1]Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, e-mail: kankova@utia.cas.cz

$\bar{g}_i(:= \bar{g}_i(x),\, x \in R^n)$, $i = 1, \ldots, s$ and $\alpha_i \in (0, 1)$, $i = 1, \ldots, , s$ such that

$$\text{either} \quad X_F(:= X_F(\alpha)) = \bigcap_{i=1}^{s} \{x \in X : P[\omega : \bar{g}_i(x) \leq \xi_i] \geq \alpha_i\},\ \alpha = (\alpha_1, \ldots, \alpha_s), \tag{3}$$

the case (3) corresponds to a special form of individual probability constraints;

$$\begin{aligned} \text{or} \quad X_F(:= X_F(u_0, \alpha)) &= \bigcap_{i=1}^{s} \{x \in X : \min_{u^i} \{P[\omega : L_i(x, \xi) \leq u^i] \geq \alpha_i\} \leq u_0^i\}, \\ &\quad u_0^i > 0,\ u_0 = (u_0^1, \ldots, u_0^s),\ \alpha = (\alpha_1, \ldots, \alpha_s), \\ L_i(x, z) &= \bar{g}_i(x) - z_i,\ i = 1, \ldots, s,\ z = (z_1, \ldots, z_s), \end{aligned} \tag{4}$$

$L_i(x, z)$, $i = 1, \ldots, s$ can be considered as loss functions. This type of loss functions can appear, e.g., in connection with an inner problem of two stage stochastic (generally nonlinear) programming problems (for definition of two–stage problems see, e.g. [1]).

To analyze the problem (1), the results of the multiobjective deterministic problems will be recalled. Since, it follows from them that the results of one–objective optimization theory can be useful to investigate the relationship between the results obtained under complete knowledge of P_F and them obtained on the data base, we recall also one–objective stochastic programming problems.

2 Some Definition and Auxiliary Assertion

2.1 Deterministic Case

To recall some results of the multiobjective deterministic optimization theory we consider a multiobjective deterministic optimization problem in the following form:

$$\text{Find} \quad \min f_i(x),\ i = 1, \ldots, l \quad \text{subject to} \quad x \in \mathcal{K}, \tag{5}$$

where $f_i(x)$, $i = 1, \ldots, l$ are real–valued functions defined on R^n, $\mathcal{K} \subset R^n$ is a nonempty set.

Definition 1. The vector x^* is an efficient solution of the problem (5) if and only if there exists no $x \in \mathcal{K}$ such that $f_i(x) \leq f_i(x^*)$ for $i = 1, \ldots, l$ and such that for at least one i_0 one has $f_{i_0}(x) < f_{i_0}(x^*)$.

Definition 2. The vector x^* is properly efficient solution of the multiobjective optimization problem (5) if and only if it is efficient and if there exists a scalar $M > 0$ such that for each i and each $x \in \mathcal{K}$ satisfying $f_i(x) < f_i(x^*)$ there exists at least one j such that $f_j(x^*) < f_j(x)$ and

$$\frac{f_i(x^*) - f_i(x)}{f_j(x) - f_j(x^*)} \leq M. \tag{6}$$

Proposition 1. ([4]) *Let $\mathcal{K} \subset R^n$ be a nonempty convex set and let $f_i(x)$, $i = 1, \ldots, l$ be convex functions on $\mathcal{K}$. Then x^0 is a properly efficient solution of the problem (5) if and only if x^0 is optimal in*

$$\min_{x \in \mathcal{K}} \sum_{i=1}^{l} \lambda_i f_i(x) \quad \text{for some} \quad \lambda_1, \ldots, \lambda_l > 0; \quad \sum_{i=1}^{l} \lambda_i = 1.$$

A relationship between efficient and properly efficient points is presented e.g in [3] or in [4].

Remark 1. Let $f(x) = (f_1(x), \ldots, f_l(x))$, $x \in \mathcal{K}$; $\mathcal{K}^{eff}$, $\mathcal{K}^{peff}$ be sets of efficient and properly efficient points of the problem (5). If $\mathcal{K}$ is a convex set, $f_i(x)$, $i = 1, \ldots, l$ are convex functions on $\mathcal{K}$, then

$$\mathcal{K}^{peff} \subset \mathcal{K}^{eff} \subset \bar{\mathcal{K}}^{peff} \quad \text{where} \quad \bar{\mathcal{K}}^{peff} \quad \text{denotes the closure set of} \quad \mathcal{K}^{peff}. \tag{7}$$

It follows from Proposition 1 that the properties of the multiobjective optimization can be (under some assumptions) investigated by one–objective theory. To recall suitable results, let $\|\cdot\| = \|\cdot\|_n$ denote the Euclidean norm in R^n and $\Delta_n[\cdot, \cdot]$ the Hausdorff distance of subsets of R^n (for the definition of the Hausdorff distance see e.g. [13]).

Proposition 2. *Let X be a nonempty compact set. If*

1. *$\hat{g}_0 := \hat{g}_0(x)$, $x \in R^n$ is a Lipschitz function on X with the Lipschitz constant L,*
2. *$\bar{X}(v) \subset R^n$ are nonempty sets for $v \in Z_F$ and, moreover, there exists $\hat{C} > 0$ such that*

$$\Delta_n[\bar{X}(v(1)),\, \bar{X}(v(2))] \leq \hat{C}\|v(1) - v(2)\| \quad \text{for} \quad v(1),\, v(2) \in Z_F,$$

then

$$|\inf_{x\in\bar{X}(v(1))} \hat{g}_0(x) - \inf_{x\in\bar{X}(v(2))} \hat{g}_0(x)| \leq L\hat{C}\,\|v(1) - v(2)\| \quad \text{for} \quad v(1),\, v(2) \in Z_F.$$

Proof. Proposition 2 is a slightly modified version of Proposition 1 in [6]. □

Remark 2. An estimate of $\hat{C}$ can be found (for some special form of $\bar{g}_i$, $i = 1, \ldots, s$) in [6].

Definition 3. Let $h(x)$ be a real–valued function defined on a nonempty convex se $\mathcal{K} \subset R^n$. $h(x)$ is a strongly convex (with a parameter $\rho > 0$) function if

$$h(\lambda x^1 + (1-\lambda)x^2) \leq \lambda h(x^1) + (1-\lambda)h(x^2) - \lambda(1-\lambda)\rho\|x^1 - x^2\|^2 \quad \text{for every} \quad x^1,\, x^2 \in \mathcal{K},\, \lambda \in \langle 0,\, 1\rangle.$$

Proposition 3. *[5] Let $\mathcal{K} \subset R^n$ be a nonempty, compact, convex set. Let, moreover, $h(x)$ be a continuous strongly convex (with a parameter $\rho > 0$) real–valued function defined on $\mathcal{K}$. If x^0 is defined by the relation $x^0 = \arg\min_{x\in\mathcal{K}} h(x)$, then*

$$\|x - x^0\|^2 \leq \frac{2}{\rho}|h(x) - h(x^0)| \quad \text{for every} \quad x \in \mathcal{K}.$$

2.2 One–Objective Stochastic Programming Problems

To recall suitable for us assertions of one criteria stochastic optimization theory we start with the problem:

$$\text{Find} \quad \varphi(F,\, X_F) = \inf \mathsf{E}_F g_0(x,\, \xi) \quad \text{subject to} \quad x \in X_F, \tag{8}$$

where $g_0(x,\, z)$ is a real–valued function defined on $R^n \times R^s$.

First, if F and G are two s–dimensional distribution functions for which the Problem (8) is well defined and if $\mathcal{X}(F,\, X_F)$ denotes a solution set of the problem (8), then we can obtain that

$$|\varphi(F,\, X_F) - \varphi(G,\, X_G)| \leq |\varphi(F,\, X_F) - \varphi(G,\, X_F)| + |\varphi(G,\, X_F) - \varphi(G,\, X_G)|. \tag{9}$$

If $g_0(x,\, z)$ is a strongly convex functions on X (with a parameter $\rho > 0$), then $\mathcal{X}(F,\, X_F)$ is singleton and consequently also

$$\|\mathcal{X}(F,\, X_F) - \mathcal{X}(G,\, X_G)\| \leq \|\mathcal{X}(F,\, X_F) - \mathcal{X}(G,\, X_F)\| + \|\mathcal{X}(G,\, X_F) - \mathcal{X}(G,\, X_G)\|. \tag{10}$$

According to (9), (10) we can study separately stability of the problem (8) with respect to perturbation in the objective function and in the constraints set. To this end let F_i, $i = 1, \ldots, s$ denote one–dimensional marginal distribution functions corresponding to F and $k_{F_i}(\alpha_i)$ be defined by the relation:

$$k_{F_i}(\alpha_i) = \sup\{z_i :\, P_{F_i}\{\omega : z_i \leq \xi_i(\omega)\} \geq \alpha_i\},\, \alpha_i \in (0,\, 1),\, i = 1, \ldots, s. \tag{11}$$

Proposition 4. *Let $\bar{g}_i$, $i = 1, \ldots, s$ be real-valued continuous function defined on R^n, $\alpha_i \in (0,\, 1)$, $i = 1, \ldots, s$, $\alpha = (\alpha_1, \ldots, \alpha_s)$. If P_{F_i}, $i = 1, \ldots, s$ are absolutely continuous with respect to the Lebesque measure on R^1, then*

$$X_F = \bar{X}(k_F(\alpha)), \quad k_F(\alpha) = (k_{F_1}(\alpha_1), \ldots, k_{F_s}(\alpha_s)),$$

where

$$\bar{X}(v) = \bigcap_{i=1}^{s} \{x \in X :\, \bar{g}_i(x) \leq v_i\},\, v = (v_1, \ldots, v_s) \quad \text{in the case} \quad (3),$$

$$\bar{X}(v) = \bigcap_{i=1}^{s} \{x \in X :\, \bar{g}_i(x) - u_0^i \leq v_i\},\, v = (v_1, \ldots, v_s) \quad \text{in the case} \quad (4).$$

Proof. The assertion of Proposition 4 has been proven in [6] (the case (3)) and in [8] (the case 4). □

Proposition 5. *Let the assumptions of Proposition 2 and Proposition 4 be fulfilled. Then*

$$|\inf_{\bar{X}(k_F(\alpha))} \hat{g}_0(x) - \inf_{\bar{X}(k_G(\alpha))} \hat{g}_0(x)| \leq L\hat{C}\|k_F(\alpha) - k_G(\alpha)\|.$$

If, moreover, $\hat{g}_0(x)$ is a strongly convex function on X, then there exists a constant $\bar{C}$ such that

$$\|\mathcal{X}(F,\, X_F) - \mathcal{X}(F,\, X_G)\|^2 \leq \bar{C}\|k_F(\alpha) - k_G(\alpha)\|.$$

Proof. The first assertion of Proposition 5 follows from Proposition 2 and Proposition 4. To prove the second assertion we consider the distribution function $\bar{F}$ fulfilling the relations

$$k_{\bar{F}}(\alpha) = (k_{\bar{F}_1}(\alpha_1), \ldots,\, k_{\bar{F}_s}(\alpha_s)), \quad k_{\bar{F}_i}(\alpha_i) = \max[k_{F_i}(\alpha_i),\, k_{G_i}(\alpha_i)],\, i = 1,\, \ldots,\, s$$

and employ the relation (10), the first assertion and Proposition 3. □

Furthermore, if we replace G by F^N we can employ the previous assertions to investigate the relationship between $\varphi(F,\, X_F)$, $\mathcal{X}(F,\, X_F)$ and $\varphi(F^N,\, X_{F^N})$, $\mathcal{X}(F^N,\, X_{F^N})$. To this end we assume:

A.2 $\{\xi^i\}_{i=1}^{\infty}$ is an independent random sequence corresponding to F; F^N is determined by $\{\xi^i\}_{i=1}^{N}$,

A.3 $P_{F_i}, i = 1, \ldots, s$ are absolutely continuous w. r. t. the Lebesgue measure on R^1,

A.4 for $i \in \{1, \ldots, s\}$ there exist $\delta > 0, \vartheta > 0$ such that $\bar{f}_i(z_i) > \vartheta \quad \text{for} \quad z_i \in Z_{F_i},\, |z_i - k_{F_i}(\alpha_i)| < 2\delta$, where $\bar{f}_i := \bar{f}_i(z_i)$, $i = 1, \ldots, s$ denotes the probability density corresponding to F_i.

Proposition 6. *[7] Let $s = 1$, $\alpha \in (0,\, 1)$. If Assumption A.2, A.3 and A.4 are fulfilled, $0 < t^{'} < \delta$, then*

$P\{\omega : |k_{F^N}(\alpha) - k_F(\alpha)| > t^{'}\} \leq 2\exp\{-2N(\vartheta t^{'})^2\}, \quad N \in \mathcal{N}, \quad (\mathcal{N}$ denotes the set of natural numbers).

Proposition 7. *Let X be a convex, compact and nonempty set, $\alpha_i \in (0,\, 1)$, $i = 1, \ldots, s$, $\alpha = (\alpha_1, \ldots, \alpha_s)$, $N \in \mathcal{N}$. If assumptions A.2, A.3, A.4 and assumptions of Proposition 5 are fulfilled, then there exists a constant $C > 0$ such that*

$$P\{\omega : |\inf_{\bar{X}(k_F(\alpha))} \hat{g}_0(x) - \inf_{\bar{X}(k_{F^N}(\alpha))} \hat{g}_0(x)|| > t\} \leq 2s\, \exp\{-2N(\vartheta t/LCs)^2\} \quad \text{for} \quad 0 < t/LCs < \delta.$$

If moreover $\hat{g}_0(x)$ is a strongly convex function on X, then there exists a constant $\bar{C}$ such that

$$P\{\omega : \|\mathcal{X}(F,\, X_F) - \mathcal{X}(F,\, X_{F^N})\|^2 > t\} \leq 2s\, \exp\{-2N(\vartheta t/L\bar{C}s)^2\}.$$

Proof. The assertion of Proposition 7 follows from Proposition 5, Proposition 6 and the properties of the Euclidean norm. □

Furthermore, we replace $g_0(x,\, z) := \hat{g}_0(x)$ independently of $z \in Z_F$ by more general case when

1a. $g_0(x,\, z)$ is a real–valued Lipschitz function on X with the Lipschitz constant $L^{'}$ not depending on $z \in Z_F$.

Then $\mathsf{E}_{F^N} g_0(x,\, \xi)$ is a Lipschitz function of $x \in X$ with the Lipschitz constant $L^{'}$ not depending on $\omega \in \Omega$. Consequently, according to Proposition 2 and Proposition 4 there exists $C^{'}$ such that

$$|\inf_{\bar{X}(k_F(\alpha))} \mathsf{E}_{F^N}\, g_0(x,\, \xi) - \inf_{\bar{X}(k_{F^N}(\alpha))} \mathsf{E}_{F^N} g_0(x,\, \xi)| \;\leq\; L^{'}C^{'}\|k_F(\alpha) - k_{F^N}(\alpha)\| \quad \text{for} \quad \omega \in \Omega$$

and for $t > 0$

$$P\{\omega : |\inf_{\bar{X}(k_F(\alpha))} \mathsf{E}_{F^N}\, g_0(x,\, \xi) - \inf_{\bar{X}(k_{F^N}(\alpha))} \mathsf{E}_{F^N} g_0(x,\, \xi)| > t\} \leq P\{\omega : L^{'}C^{'}\|k_{F^N}(\alpha) - k_F(\alpha)\| \geq t\}.$$

If, moreover, $g_0(x,\, z)$ is a strongly convex (with a parameter $\rho > 0$) function on X, then employing the proof technique from the proof of Proposition 5 we can obtan that

$$P\{\omega : \|\mathcal{X}(F,\, X_F) - \mathcal{X}(F,\, X_{F^N})\|^2 > t\} \leq P\{\omega : 2L^{'}C^{'}\rho\|k_{F^N}(\alpha) - k_F(\alpha)\| \geq t\}$$

Furthermore, employing Proposition 7 and the relations (9), (10) we obtain

Proposition 8. *Let the assumptions A.2, A.3, A.4 and 1.a are fulfilled, $t > 0$. If there exists $\gamma \in (0,\, 1/2)$ such that*

$$P\{\omega : N^{\gamma}|\varphi(F,\, X_F) - \varphi(F^N,\, X_F)| > t\} \xrightarrow[N\to\infty]{} 0,$$

then

$$P\{\omega : N^{\gamma}|\inf_{\bar{X}(k_F(\alpha))} \mathsf{E}_F\, g_0(x,\, \xi) - \inf_{\bar{X}(k_{F^N}(\alpha))} \mathsf{E}_{F^N} g_0(x,\, \xi)| > t\} \xrightarrow[N\to\infty]{} 0.$$

Moreover, if $g_0(x,\, z)$ is for every $z \in Z_F$ a strongly convex function of $x \in X$ with a parameter $\rho >, 0$ then also

$$P\{\omega : N^{\gamma}\|\mathcal{X}(F,\, X_F) - \mathcal{X}(F^N,\, X_{F^N})\|^2 > t\} \xrightarrow[N\to\infty]{} 0.$$

3 Multiobjective Stochastic Case

To analyze stochastic multiobjective problem (1) we restrict our consideration to the case when

B.1 $g_i(x,\, z)$, $i = 1,\, \ldots,\, l$ are for every $z \in Z_F$ strongly (with a parameter $\rho > 0$) convex functions; moreover they are Lipschitz on X with the Lipschitz constant L_1 not depending on $z \in Z_F$.

Furthermore we introduce (for two s–dimensional distribution functions F, G) the sets $\mathcal{G}(F,\, X_F)$, $\bar{\mathcal{X}}(F,\, X_F)$, $\bar{\mathcal{G}}(F,\, X_F)$, Λ and the function $\bar{g}(x,\, z,\, \lambda)$ by the relations:

$$\begin{aligned}
\mathcal{G}(F,\, X) &= \{y \in R^l : y_j = \mathsf{E}_F g_j(x,\, \xi),\, j = 1,\, \ldots,\, l \quad \text{for some} \quad x \in X;\, y = (y_1,\, \ldots,\, y_l)\},\\
\bar{\mathcal{X}}(F,\, X_F) &= \{x \in X_F : x \quad \text{is a properly efficient point of the problem (1)}\},\\
\bar{\mathcal{G}}^F(F,\, X_F) &= \{y \in R^l :\, y_j = \mathsf{E}_F g_j(x,\, \xi),\, j = 1,\, \ldots,\, l \quad \text{for some} \quad x \in \bar{\mathcal{X}}(F,\, X_F)\},\\
\bar{\mathcal{G}}^F(G,\, X_G) &= \{y \in R^l :\, y_j = \mathsf{E}_F g_j(x,\, \xi),\, j = 1,\, \ldots,\, l \quad \text{for some} \quad x \in \bar{\mathcal{X}}(G,\, X_G)\},\\
\Lambda &= \{\lambda \in R^l :\, \lambda = (\lambda_1,\, \ldots,\, \lambda_l),\, \lambda_i > 0,\, i = 1,\, \ldots,\, l,\, \sum_{i=1}^{l} \lambda_i = 1\},\\
\bar{g}(x,\, z,\, \lambda) &= \sum_{i=1}^{l} \lambda_i g_i(x,\, z), \quad x \in R^n,\, z \in R^s,\, \lambda \in \Lambda.
\end{aligned} \tag{12}$$

Evidently, under the assumption B.1, the function $\bar{g}(x,\, z,\, \lambda)$ is (for every $z \in Z_F$, $\lambda \in \Lambda$) a (strongly with a parameter $\rho > 0$) convex and Lipschitz function on X with the Lipschitz constant L_1 not depending on $z \in Z_F$. Consequently, if we define the parametric optimization problem

$$\text{Find} \quad \varphi^{\lambda}(F,\, X_F) = \inf \mathsf{E}_F \bar{g}(x,\, \xi,\, \lambda)) \quad \text{subject to} \quad x \in X_F \quad \text{for} \quad \lambda \in \Lambda, \tag{13}$$

we can employ the auxiliary assertions of previous parts to analyze the Problem (1) and to see that the following Theorems are valid.

Theorem 9. *Let Assumptions B.1, A.2, A.3 and A.4 be fulfilled, X be a compact nonempty convex set, $g_i(x,\, z) := g_i^0(x)$, $x \in X$, $z \in Z_F$, $i = 1,\, \ldots,\, l$, then*

$$\begin{aligned}
P\{\omega : \Delta_l[\mathcal{G}(F,\, X_F),\, \mathcal{G}(F^N,\, X_{F^N})] \longrightarrow_{N\longrightarrow\infty} 0\} &= 1,\\
P\{\omega : \Delta_n[\bar{\mathcal{X}}(F,\, X_F),\, \bar{\mathcal{X}}(F^N,\, X_{F^N})] \longrightarrow_{N\longrightarrow\infty} 0\} &= 1,\\
P\{\omega : \Delta_l[\bar{\mathcal{G}}^F(F,\, X),\, \bar{\mathcal{G}}^F(F^N,\, X_{F^N})] \longrightarrow_{N\longrightarrow\infty} 0\} &= 1.
\end{aligned}$$

Remark 3. It follows (under the assumptions of Theorems 9) that asymptotic properties, in the case of constraints sets fulfilling the relation (3), do not depend on the tails of distributions (including all stable case).

Theorem 10. *Let B.1, A.2, A.3 be fulfilled, X be a compact nonempty set, $t > 0$. If there exists $\gamma \in (0,\, 1/2)$ such that*

$$P\{\omega : N^{\gamma}|\inf_{X_F(\alpha)} \mathsf{E}_F g_i(x,\, \xi) - \inf_{X_F(\alpha)} \mathsf{E}_{F^N} g_i(x,\, \xi)| > t\} \longrightarrow_{N\longrightarrow\infty} = 0, \tag{14}$$

then

$$\begin{aligned} P\{\omega : N^{\gamma}\,\Delta_l[\mathcal{G}(F,\,X),\,\mathcal{G}(F^N,\,X_{F^N})] > t\} \quad &\longrightarrow_{N\longrightarrow\infty} \quad = \quad 0. \\ P\{\omega : N^{\gamma/2}\,\Delta_n[\bar{\mathcal{X}}(F,\,X),\,\bar{\mathcal{X}}(F^N,\,X_{F^N})] > t\} \quad &\longrightarrow_{N\longrightarrow\infty} \quad = \quad 0, \\ P\{\omega : N^{\gamma/2}\,\Delta_l[\bar{\mathcal{G}}^F(F,\,X),\,\bar{\mathcal{G}}^F(F^N,\,X_{F^N})] > t\} \quad &\longrightarrow_{N\longrightarrow\infty} \quad = \quad 0. \end{aligned} \tag{15}$$

If, moreover, there exist functions $\hat{g}_i^0(x)$, $\imath = 1, \ldots, l$ *defined on* R^n *such that* $g_i(x,\,z) = \hat{g}_i^0(x)$, $i = 1, \ldots,$ *then the relations (15) are valid with* $\gamma \in (0,\,1/2)$.

Remark 4. It follows from Theorems 9, 10 that asymptotic properties, in the case of constraints sets fulfilling the relation (3) or (4) and the condition $g_i(x,\,z) := g_0^i(x,\,z)$, $i = 1, \ldots, l$ (fulfilling the condition 1.a) do not depend on the tails of distributions (including stable distributions). To this end it is assumed that the one-dimensional marginals are absolutely continuous with respect to Lebesque measure in R^1. Practically this assertion follows from the old results of [2] or [14]. The distribution with heavy tails appear very often in economic and financial applications, see e.g. [11], [12].

4 Conclusion

The paper deals with multiobjective stochastic programming problems, especially with a relationship between characteristics of these problems corresponding to complete knowledge of the probability measure and those determined on the data base. However, in spite of the former papers there the assumption of "deterministic" constraints set is rather relaxed. On the other hand we still assume that the objective functions are strongly convex with respect to $x \in X$. Evidently linear functions do not fulfil this assumptions. It seems (under the corresponding analysis) that similar assertions are also valid in the linear case. However, a detailed investigation in this direction is beyond the scope of this paper.

Acknowledgment This work was supported by the Czech Science Foundation under grant 13-14445S.

References

[1] Birge, J.R. and Louveaux, F.: *Introduction in Stochastic Programming.* Springer, Berlin 1999.

[2] Dvoretski, A., Kiefer, J., and Wolfowitz, J.: Asymptotic minimax character of the sample distribution function and the classical. *Ann. Math. Statist.* **27** (1956),642–669.

[3] Ehrgott, M: *Multicriteria Optimization.* Springer, Berlin, 2005.

[4] Geoffrion, A. M.: Proper efficiency and the theory of vector maximization. *J. Math. Annal. Appl.* **22** (1968), 3, 618–630.

[5] V. Kaňková, V.: Stability in stochastic programming – the case of unknown location parameter. *Kybernetika* **29** (1993), 1, 97–112.

[6] Kaňková, V.: On the stability in stochastic programming: the case of individual probability constraints. *Kybernetika* **33** (1997), 5, 525–544.

[7] Kaňková, V.: Empirical estimates in optimization problems: survey with special regard to heavy tails and dependent samples. *Bulletin of the Czech Econometric Society* **19** (2012), 30, 92–111.

[8] Kaňková, V.: Risk measures in optimization problems via empirical estimates. *Czech Economic Review, Acta Universitatis Carolinea* **7** (2012), 3, 162–177.

[9] Kaňková, V.: Economic and financial problems via multiobjective stochastic optimization. *Proceedings of the 31st International Conference Mathematical Methods in Economics 2013* (Jana Vojáčková, ed.), College of Polytechnics Jihlava (Czech Republic) 2013, 362–367.

[10] Kaňková, V.: A remark on multiobjective stochastic optimization in economic and financial problems. *Central European Journal of Operations Research* (submitted).

[11] Kozubowski, T.J., Panorska A. K., and Rachev, S. T.: Statistical issues in modelling stable portfolios. In: *Handbook of Heavy Tailed Distributions in Finance* (S.T. Rachev, ed.), Elsevier, Amsterdam 2003, 134–168.

[12] Meerschaert, M.M., and Scheffler, H.P.: Portfolio modelling with heavy tailed random vectors. In: *Handbook of Heavy Tailed Distributions in Finance* (S.T. Rachev, ed.), Elsevier, Amsterdam 2003, 595–640.

[13] Rockafellar, R.T., and Wets, R.J.B.: *Variational Analysis.* Springer, Berlin, 1983.

[14] Serfling, J.R.: *Approximations Theorems of Mathematical Statistics.* Wiley, New York, 1980.

Credit and Business Cycle Co-movements in V4 countries: Evidence from Wavelet Analysis

Svatopluk Kapounek[1], Jitka Poměnková[2]

Abstract. The global shocks caused by economic crisis transmitted to liquidity and global capital flows. Reversely, the worldwide slowdown of credit flows affected investment activity and business cycle. Quantifying this transmission is one of the most important topics of the recent financial crisis, especially in the context of European recovery, because the banking system (credit supply) plays a crucial role for sustained recovery in the Europe after the financial crises.
We focus on credit and business cycles co-movements in V4 countries. The relation between the time series represents transmission of credit shocks to aggregate economic activity movements. The most of the empirical analysis focus on time domain. Our contribution is in the application of analysis in time-frequency domain. The results provide distinguish cyclical movements in different frequency band to understand co-movements in short-term and long-run.

Keywords: time-frequency domain, financial crisis, banking system

JEL Classification: E32 ; E51; C14

1 Introduction

The economic crisis has spread from financial markets to real economies in countries around the world. There is a large literature on the global transmission of past financial crises which mostly finds strong evidence for the transmission of global shocks to liquidity and global capital flows (e.g. Brunnermeier 2009; Calvo 2009 or Kalemli-Ozcan, Papaioannou and Perri, 2010). The studies have resulted in a worldwide slowdown of credit flows which affected investment activities and business cycle. The effect of changes in bank capital on the banks' lending is a key determinant of the linkage between financial conditions and real activity. Quantifying of this transmission is one of the most important topics of the recent financial crisis, especially in the context of European recovery. The empirical analysis of 16 emerging European countries was provided by Popov and Udell (2012). They analysed sensitivity of credit supply to bank financial condition and found that decline in banking equity, Tier 1 capital and losses on financial assets reduced the credit flows to the firms during the crisis.

The tightening of the conditions required to obtain a banking credits increases the dangers of the liquidity squeeze becoming a credit crunch. A popular view seems to be that this decline in investment activities is driven by a credit crunch through a financial accelerator effect (e.g. Bernanke et al., 1996; Fidrmuc et al., 2010 or Korinek, 2011). Several studies have investigated different determinants of credit sources availability. Geršl and Jakubík (2009) or Memmerl, Schmieder and Stein (2007) analyzed models of bank financing and its effects on credit risk of the banks and credit availability. Jakubík and Teplý (2008) focused on scoring and its impact on creditworthiness of the Czech corporate sector.

Conversely, Peek and Rosengren (1995) argue that credit availability reflects normal procyclical pattern of both the creditworthiness of borrowers and credit demand. The use of credit by firms is explained by several theories. The financing theory suggests that firms with bank credits can offer higher trade credits to their customers to increase competitiveness at the market (Deloof and Van Overfelt, 2008). The financing advantage theory and transaction theory explain credit use by both suppliers and buyers (Schwartz, 1974, or Summers and Wilson, 2002). The marketing theory provides an alternative approach. The motivation of using trade credit creates long-term relationship with customers and ensures long-term benefits through future sales to these customers (Summers and Wilson, 2002).

The emphasis of past work has been mostly focused on empirical methods to identify the factors and their effects on slowdown of credit flows. In this paper we follow money endogeneity hypothesis, especially the causality where money are created by the credit creation determined by the demand factors. This assumption was con-

[1] Faculty of Business and Economics, Mendel University in Brno, Department of Finance, Zemedelska 1, 613 00 Brno, email: kapounek@mendelu.cz
[2] Faculty of Electrical Engineering and Communication, Brno University of Technology, Dept.of Radio Electronics, Technická 12, 616 00 Brno, pomenkaj@feec.vutbr.cz

firmed by many empirical and theoretical studies after the financial crisis and break down of the economic activity in the all euro are member states (Setzer and Wolff, 2009 or Pomenkova and Kapounek, 2013). Summarily we suppose that the direction of the causality is from business to credit cycle.

The other part of the recent literature deals with credit and business cycles co-movements. The relation between the time series represents transmission of credit shocks to aggregate economic activity movements. The most of the empirical analysis focus on time domain (Kocherlakota, 2000, Xu, 2012 or Bahadir and Gumus, 2012). The analysis in frequency domain provides distinguish cyclical movements in different frequency band to understand co-movements in short-term and long-run (Zhu, 2010). However, the application of analysis in frequency, especially time-frequency domain is unresearched.

The objective of the paper is to provide analysis of business and credit cycle co-moements in time-frequency domain and compare the results before and after the financial crisis. Finally, we discuss co-movements in time-frequency domain as the alternate methodological approach to identify disequilibrium at the money market

2 Methods

Measuring of synchronization between economic time series in the time domain predominated for long time. Widely use technique is correlation analysis or its extension such as the cointegration, common features (Engle and Kozicki, 1993) or common cycles and codependence (Vahid and Engle, 1997). Unfortunately analysis in time domain does not provide in deep view into the time series structure. Therefore we do not have sufficient information to answer the question how to quantify the degree of synchronisation and how to analyse the evolution of such a synchronisation during time with respect to the occurrences in the economic time series.

In the past several decades methodological approaches proceeded to techniques in frequency domain (spectral and cross spectral analysis) which allow to viewed the comovement between time series from the perspectives of a dynamic correlation and phase shift methods, coherence or a squared coherency. The theoretical background with a practical application on business cycles in Europe and the USA provide Croux et al. (2001). The practically oriented studies currently use dynamic correlation (Fidrmuc et al., 2012, Poměnková et al., 2014) in frequency domain. Interconnection of both time and frequency domain provides time-frequency approaches such as a wavelet analysis, a time-varying AR process or a time-varying periodogram (Blumenstain and all, 2011). The combination of both domains provides a more efficient means of statistical analysis. Aguiar-Conraria and Soares (2011) use a wavelet analysis to study business cycle synchronisation across the euro countries. Poměnková et al. (2014) use wavelet spectrum analysis to study globalization and business cycles in China and G7 countries. The advantage of the wavelet analysis (wavelet power spectrum, the wavelet cross-spectrum and to wavelet coherence) is that it can capture the features of non-stationarity time series due to the simultaneous time-frequency decomposition of inputs (Jiang and Mahadevan, 2011).

The continuous wavelet transform (CWT) of time series *s(t)* is usually defined as the following integral (Gencay et al., 2002) as

$$S_{CWT}(a,b) = \frac{1}{\sqrt{b}} \int_{-\infty}^{\infty} s(t)\psi\left(\frac{t-a}{b}\right)dt, \quad b > 0, a \in R, \tag{1}$$

where *a* denotes the translation (time shift) and *b* denotes dilatation (scale) of the mother wavelet $\psi(\cdot)$. The dilatation is related to the Fourier frequency and numerator of the fraction $\sqrt{b}$ ensures the conservation of energy. The CWT transforms input time series from the time domain representation to the time-scale domain.

In order to identify comovement between two time series we can use measures via several quantities (Croux et al., 2001). In the frequency domain we can use dynamic correlation, the coherence or cohesion. Similar analysis could be also done in the time-scale domain. In our approach we use the wavelet cross spectrum between two inputs $s_1(t)$ and $s_2(t)$ for their time-scale representation $S_{CWT\text{-}1}(a,b)$ and $S_{CWT\text{-}2}(a,b)$. We follow definition presented by Jiang and Mahadevan (2011) which can be defined as

$$S_{12} = S_{CWT-1}(a,b)S_{CWT-2}(a,b). \tag{3}$$

3 Data

We use non-seasonally adjusted monthly data of loans (ECB, balance sheets of monetary financial institutions) and industrial production index (Eurostat, short-term business statistics) of the Czech Republic (CZ) in 2002/01-2014/02, Hungary (HU) in 2003/01-2014/02, Poland (PL) 2004/01-2014/02, Slovakia (SK) in 2006/01-2014/02 and Germany (DE) in 2003/01-2014/02. Time series are in prices of the year 2005. The analysis is performed on two data sets: at first on whole data set and at the second at the shortened data set 2009/01-2014/02 (available

sample is n=62). The selection of the time 2009 were motivated by the financial crisis. The data were transformed by natural logarithm and pre-filtered by Hodrick-Prescott filter (Hodrick and Prescott, 1980). As stated in Christiano and Fitzgerald (2003) if the raw data before application of band-pass filters have a non-zero mean or are covariance stationary about a trend, then the trend has to be removed prior to analysis of optimal filter design.

4 Results

For evaluation of co-movement we use wavelet co-spectrum and Morlet mother wavelet functions (Gencay et al. 2002). The results are presented in three types of charts. The first type of figures (1a-5a) presents wavelet co-spectrum between loans and industrial production index for each country in whole available time period. The second type of figures (1b-5b) presents the same result but on the shortened samples, i.e. from starting time till the start of financial crisis in 2008. The third type of figure (1c-5c) presents wavelet co-spectrum after 2009 till the end of available data set. As we can see on figures 1a-4a the highest level of the values of the cospectrum were achieved between years 2008-2009 in all countries. It caused that other frequency component are not such visible. Therefore we decided to split up time period with respect to the time. The cospectrum is measured on the y-axis for specific periods (x-axis) and periodicities (z-axis). The intensity of each contour represents the relative importance of the different periodicities and time.

In the all countries we can identify evident impact of the financial crisis represented by significant co-movement of credit and business cycle (the peak in the Figure 1a-5a around year 2008). However, frequency differentiation provides much more interesting results. We applied non-seasonally adjusted time series. Thus, in comparison to our expectations there is no evidence of seasonal co-movement. It means that seasonal sectors (e.g. building and construction).

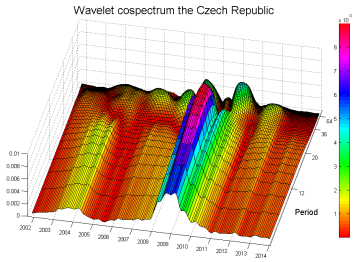


Figure 1a Wavelet co-spectrum CZ, 2002/01-2014/02

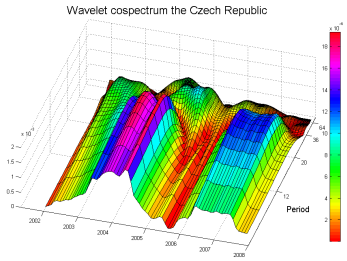


Figure 1b Wavelet co-spectrum CZ, 2002/01-2007/12

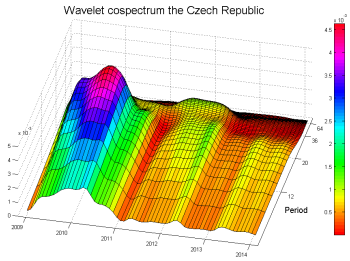


Figure 1c Wavelet co-spectrum CZ, 2009/01-2014/02

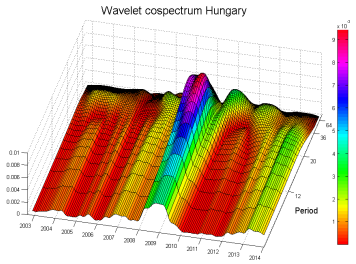


Figure 2a Wavelet co-spectrum HU, 2003/01-2014/02

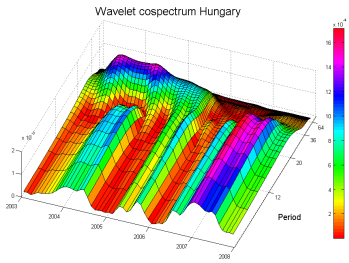


Figure 2b Wavelet co-spectrum HU, 2003/01-2007/12

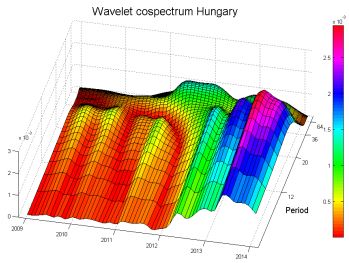


Figure 2c Wavelet co-spectrum HU, 2009/01-2014/02

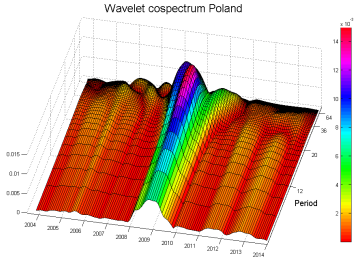


Figure 3a Wavelet co-spectrum PL, 2004/01-2014/02

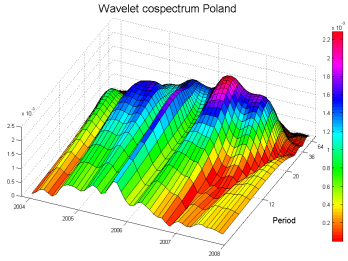


Figure 3b Wavelet co-spectrum PL, 2004/01-2007/12

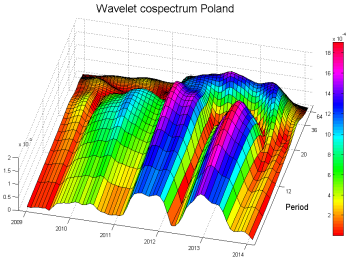


Figure 3c Wavelet co-spectrum PL, 2009/01-2014/02

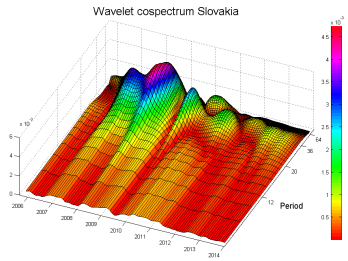


Figure 4a Wavelet co-spectrum SK, 2006/01-2014/02

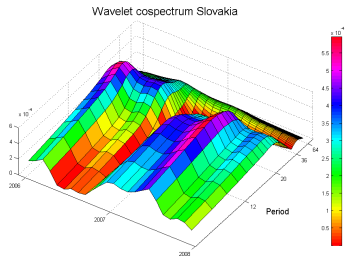


Figure 4b Wavelet co-spectrum SK, 2006/01-2007/12

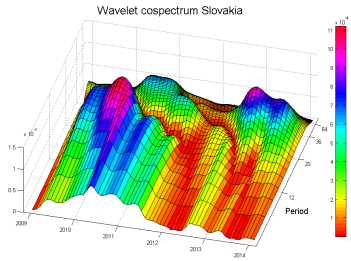


Figure 4c Wavelet co-spectrum PL, 2009/01-2014/02

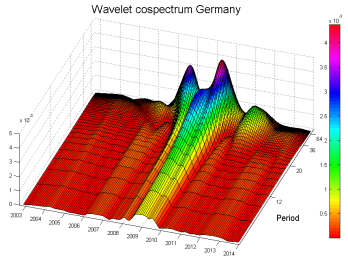


Figure 5a Wavelet co-spectrum DE, 2003/01-2014/02

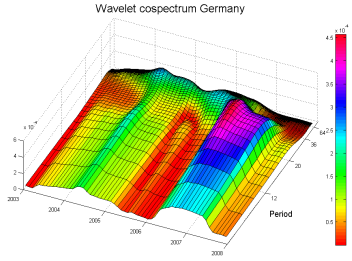


Figure 5b Wavelet co-spectrum DE, 2003/01-2007/12

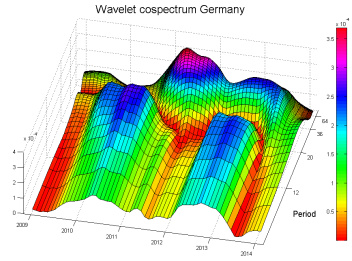


Figure 5c Wavelet co-spectrum DE, 2009/01-2014/02

In the case of the Czech Republic we can identify additional significant co-movement between the years 2003-2004 and weaker co-movement just before the financial crisis in the years 20066-2007. These co-movements before the crisis were identified in the short and medium term below 7 years. We can assume that this co-movement could be caused by the economic growth followed by credit money creation by the commercial banks. After the financial crisis we did not identify co-movements.

The Hungarian business and credit cycles were co-moved in the years in long-run during the years 2003-2005. In comparison to the Czech Republic we identified co-movement also after the financial crisis, especially after the year 2012. The development of the Poland is similar to the Hungary. The pre-crisis co-movement was identified after the year 2005. Contrary to the Hungary the post-crisis co-movement is divided into the two time periods which follow financial crisis consequences and debt-crisis development in the Europe.

The co-movement development in the Slovakia is totally different in comparison with the other discussed countries. Pre-crisis period is typical by co-movement in low frequencies (long periods over 5 years). After the financial crisis we identified twice co-movement which differs in frequencies. The first, caused directly by the crisis are evident over the all frequencies. The second post-crisis co-movement which could be caused by the subsequent debt-crisis is significant only at the low frequencies (over 4 years).

Finally, we provide the analysis of the co-movements for the Germany. The German economy is typical by the co-movement after the year 2004 caused by economic downturn. The development after the crisis is similar to the Slovakia and Poland there are two significant co-movements, in the years 2010-2011 and after the year 2012.

5 Conclusions

It is important to mention that the European countries are more reliant on bank credits and bank intermediation of savings, than the United States and rest of the world. Therefore, the banking system plays a crucial role for sustained recovery in the Europe after the financial crises. The key question from a policy perspective is whether the slowdown in credit flows is driven by supply (credit crunch hypothesis) or demand for credits. A credit crunch generally involves a reduction in the availability of credit independent of a rise in the official interest rates. Therefore, the solution might be aimed at changes in prudential policy (national regulatory framework of the banking system). If the slowdown in credit flows is mostly driven by credit demand and economic activity, then fiscal and monetary policy interactions aimed at expanding aggregate demand might be effective instrument.

Our results do not provide strong evidence on the existence of a “causal” relationship between output and credit but we conclude that there is a strong evidence of co-movements between business and credit cycle during the financial crisis. However, we follow the money endogeneity hypothesis, especially the thesis that national bank-

ing systems create credit money which is determined by investment and economic activity. Thus, we assume that the credit cycle is affected by business cycle through the money demand factors.

Finally, we can conclude that the identified co-movements are heterogeneous across the frequencies. It means the impact of the pre-crisis economic growth, financial crisis and debt-crisis differ across the V4 countries in terms of the possible causalities between the credit and business cycles and credit money creation as well. The policy implications of these results could be found in the common monetary policy and its efficiency after the future planned joining the Eurozone.

Acknowledgements

The results introduced in the paper have been funded with the support from the Czech Science Foundation via grant No. P403/14-28848S "Financial Crisis, Depreciation and Credit Crunch in CEECs" and by the Brno University of Technology Internal Grant Agency under project no. FEKT-S-14-2177 (PEKOS). The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation.

References

[1] Aguiar-Conraria, L., Soares, M. J.: Business cycle synchronization and the Euro: A wavelet analysis. *Journal of Macroeconomics* **33** (2011), 477–489.

[2] Bahadir, B., Gumus, I.: Credit Decomposition and Business Cycles. Working Paper presented at the 20th Annual Symposium of Society for Nonlinear Dynamics and Econometrics. (2012)

[3] Bernanke, B., Gertler, M., Gilchrist, S.: The financial accelerator and flight to quality. *Review of Economics and Statistics*, 78 (1996), 1-15.

[4] Blumenstein, J., Poměnková, J., Maršálek, R.: Comparative Study Of Time- Frequency Analysis Approaches With Application To Economic Indicators. In: *Proceedings of the 26th European Conference on Modelling and Simulation ECMS 2012* , Univ Koblenz Landau, Koblenz, Germany, 2012, 291-297.

[5] Brunnermeier, M.: Deciphering the Liqudity and Credit Crunch 2007-8. *Journal of Economic Perspectives* 23 (2009), 77-100.

[6] Calvo, G.: Financial Crises and Liquidity Shocks: A Bank-Run Perspective. NBER Working Paper 15425 (2009).

[7] Croux, C., Forni, M., and Reichlin, L.: A Measure of Comovement for Economic Variables: Theory and Empirics. *The Review of Economics and Statistics* **2** (2001), 232–241.

[8] Deloof, M., Van Overfelt, W.: Were Modern Capital Structure Theories Valid in Belgium Before World War I? *Journal of Business Finance & Accounting*, 35 (2008), 491–515.

[9] Engle, R. F, Kozicki, S.: Testing for common features. Journal of Business & Economic Statistics 11 (1993), 369-380

[10] Fidrmuc, J., Horváth, R., Horváthová, E.: Corporate Interest Rates and the Financial Accelerator in the Czech Republic. *Emerging Markets Finance & Trade*, 46, 4 (2010), 41–54.

[11] Fidrmuc, J., Ikeda, T., and Iwatsubo, K.: International transmission of business cycles: Evidence from dynamic correlations. *Economics Letters* **114** (2012), 252-255.

[12] Gençay, R., Selçuk, F., Whitcher, B.: *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press, California, 2002.

[13] Geršl, A., Jakubík, P.: Models of bank financing of Czech corporations and credit risk, Financial Stability Report 2008/2009, Czech National Bank (2009), 92–101.

[14] Hodrick, R.J., Prescott, E.C.: Post-war U.S. Business Cycles: An Empirical Investigation, mimeo, Carnegie-Mellon University, Pitsburgh, PA, (1980).

[15] Jakubík, P., Teplý, P.: Scoring as an Indicator of Financial Stability. Financial Stability Report 2007, Czech National Bank, (2008), 76-85.

[16] Jiang, W., Mahadevan, S.: Wavelet spectrum analysis approach to model validation of dynamic systems. *Mechanical Systems and Signal Processing* **25** (2011), 575–590.

[17] Kalemli-Ozcan, S., Papaioannou, E., Perri, F.: This Time is Different: Financial Integration and the 2007 Crisis. Working paper presented at the Joint conference of the European Central Bank and the Journal of International Economics, September 9.-10, 2010.

[18] Kocherlakota, N., R.: Creating Business Cycles through Credit Constraints. Federal Reserve Bank of Minneapolis Quarterly Review, 24 (2000), 2–10.

[19] Korinek, A.: Systemic risk-taking: Amplification effects, externalities, and regulatory responses. *ECB Working Paper*, 1345 (2011)

[20] Memmel,C., Schmieder, C., Stein, I.: Relationship lending – empirical evidence for Gemany. *Detsche Bundesbank Discussion Paper*, 14 (2007).

[21] Peek, J., Rosengren, E.S.: Banks and the Availability of Small Business Loans. *Federal Reserve Bank of Boston Working Paper*, 95-1 (1995).
[22] Poměnková, J., Fidrmuc, J., Korhonen, I.: China and the World economy: Wavelet spectrum analysis of business cycles, *BOFIT Discussion Papers* **5/2014**, ISBN 978-952-6699-68-4
[23] Poměnková, J., Kapounek, S.: Heterogeneous distribution of money supply across the euro area. In Tvrdon, M; Majerova, I.: Proceedings of the 10th International Scientific Conference: Economic Policy in the European Union Member Countries, (2013), 268-279
[24] Poměnková, J., Kapounek, S., Maršálek, R.: Variability of Dynamic Correlation – the Evidence of Sector-specific Shocks in V4 Countries. Forthcoming in *Prague Economic Papers*, (2014).
[25] Popov, A., Udell, G., F.: Cross-border banking, credit access, and the financial crisis. Journal of International Economics, 87 (2012), 147–161.
[26] Schwartz, R., A.: An economic model of trade credit. *Journal of financial and quantitative analysis.* 9 (1974), 643-657.
[27] Setzer, R., Wolff, G.B.: Money demand in the euro area: new insights from disaggregated data. DG ECFIN Economic Papers, 373 (2009).
[28] Summers, B., Wilson, N.: An Empirical Investigation of Trade Credit Demand. *International Journal of the Economics of Business.* 9 (2002), 257-270.
[29] Vahid, F., Engle, R.F., 1997: Codependent cycles. *Journal of Econometrics* **80**, (1997), 199-221.
[30] Xu, T., T.: The role of credit in international business cycles. Cambridge Working Papers in Economics 1202, Faculty of Economics, University of Cambridge. (2012)
[31] Zhu, F.: Credit and business cycles: some stylised facts. BIS Working Paper, August 2010.

Statistical Method and Neural Network for Sustainability Evaluation

Edward Kasem, Oldřich Trenz, Jiří Hřebíček.[1]

Abstract. A sustainability report is an organizational report that conveys the information about economic, environmental, social and governance indicators in a way that is comparable with financial reporting. There are different methods for measuring corporate performance and assessing their sustainability.

In this paper, two methods of environmental, social, governance and economic indicators assessment are described. The first one is Kren's method, which supposes that all indicators and dimensions similarly affect the assessing of the enterprises' performance, whereas the second one supposes that different indicators and dimensions don't equally affect the enterprises sustainability. The main objective of these methods is improving the benchmark between the enterprises' and providing an active participation in decision-making.

Keywords: performance indicators, statistical analysis, sustainability, neural network.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Sustainability reporting is a mainstream business activity that demonstrates the link between the organization strategy and commitment to a sustainable global economy. It engages companies in disclosure of their overall economic, environmental, social, and governance (ESG) impacts and efforts [15]. At present, this is valid primarily for large companies. They have indeed gotten involved in sustainability reporting practices. Small and medium sized enterprises (SMEs), despite their significant impacts European economy, employment and environment, show low level of engagement. Because of the important effects of SMEs that impact on the European economy, small and medium enterprises need a method or procedure to measure, control and improve their performance.

Many frameworks are used for measuring corporate performance and sustainability. Recently, the most widely used framework for sustainability reporting is Global Reporting Initiatives (GRI) [4, 5]. additionally, there are many international standards for implementing and certifying the sustainability pillars like ISO 9000, ISO 14000, ISO 18000 [12], ISO 26000 [11] and EMAS [2]. There are also many other popular specialized sustainability reporting frameworks like Sustainability Assessment of Food and Agriculture (SAFA) [25], in which Food and Agriculture reports are described.

The main goals of this paper can be summarized in two points. Firstly, two common methods which are used in sustainability assessment in the Czech Republic, are described. In the first one "Kren's method" [21], all the indicators similarly affect the assessing of the enterprises performance. Whereas in the second one "Statistical method" [15], each indicator affects the enterprise's rate (score) in a different manner. Secondly, new method for sustainability assessment using neural networks is proposed.

The paper is structured as follow. Two previous mentioned methods are described in Section 2. The steps followed in each method to achieve overall enterprise sustainability are presented. Section 3 introduces the proposed neural method for sustainability assessment. Finally, Section 4 concludes the paper.

[1]Mendel University in Brno,Faculty of Business and Economics, Department of Informatics, Zemědělská 1, Brno, Czech Republic, {edward.kasem, oldrich.trenz, jiri.hrebicek} @mendelu.cz

2 Methods for sustainability reporting

In this section, two methods used for sustainability reporting in the Czech Republic, are described.

2.1 Křen's method

One of the methods used for sustainability performance assessing in agriculture farms is Křen's method [21]. This method is proposed for assessing individual corporate performance indicators using the range of $< 0, 1 >$ where 0 represents undesirable value of the sustainability, whereas 1 represents the best value. These sets of indicators are presented in Table 1 [21]. In this method, the assessment is done according to the proposed functions which determine the appropriate level of corporate sustainability of agriculture farm. After assessing the indicators and calculating their rates (scores), the index of each sustainability

Key dimensions	Performance Indicator	Scale and (Unit)
Environmental indicators	Water and Air	Nitrogen balance
	Used Resources	Phosphorus balance
		Potassium balance
		Balance of organic materials
		Energy consumption
	Biodiversity	Intensity of plant protection
		Soil erosion
		Potential biodiversity
Economy indicators	stability	Enterprises sales
		Profit
		Indebtedness
	liquidity	Contribution to payment
		liquidity
	Rentability	Rentability
Social indicators	Labor and Employment	Remuneration
		Working hours
		Education
		Health and safety
		Employee Engagement
	Social engagement	Participation in society indicators rate:
		1-Social engagement
		2-Public Relations
		3-Involvement in the region

Table 1 ESG indicators for agriculture enterprises

dimension (economic, environmental, social and governance) should be calculated. These calculations are done by aggregating the rates (scores) of individual indicators, which belong to the same dimension. Then the average should be computed. The sustainability threshold for each dimension should be more than 0.75. The overall index is expressed as an average of the three dimensions' indexes according to the equation (1).

$$I_c = \frac{I_e + I_a + I_s}{3} \tag{1}$$

where I_c is the overall sustainability index for environmental, social and economic dimension. I_e is the sustainability index of economic dimension. I_a is the sustainability index of environmental dimension. I_s is the sustainability index of social dimension.

Křen's method illustration of the overall sustainability performance of agriculture organization (farm, company, enterprise) is shown in Figure 1 [21]. The visualization of the Křen sustainability performance

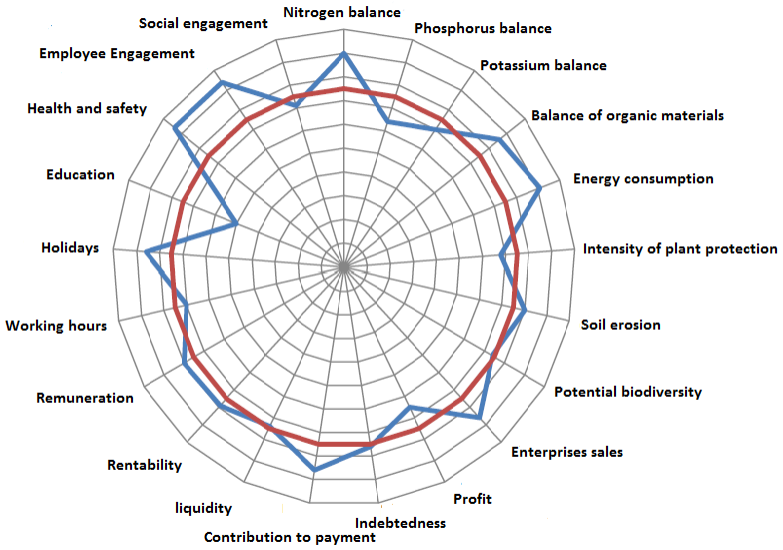


Figure 1 The result of enterprise sustainability assessment [21].

rating is depicted in the polygon of a hypothetical organization. It shows us the indicators which meet the sustainability criteria and which should be improved to achieve the completeness and balance of the agricultural system. It also gives us a clear graphical representation about the weak points in the management of the whole agricultural enterprise or its individual parts and the issues that should be included as a solution. The thick blue line connects the indicator rates for the enterprise under consideration, whereas the red one describes the boundaries of good overall sustainability.

2.2 Statistical method

The statistical method is the newest method used for sustainability assessment [15]. These methods use ESG indicators which have been developed last few years [1, 8, 9, 15] by choosing the most suitable indicator from GRI and SAFA frameworks. The sets of indicators proposed for manufacturing enterprises are listed in the tables [15, 16, 17, 18]. These indicators can be used for wide range of SMEs in the food and agriculture sector [3, 8, 10].

In this method, the sustainability indicators should be shown to the respondents of studied companies. The responses should be collected depending on Likert-type scale which is mentioned in four points: 1=not important, 2=somewhat important, 3=important, 4=very important and additional one 5=do not know. After that, several statistical analyses should be applied to all the responses that have been obtained. Depending on the statistical analysis, the weight of each indicator can be determined. This can be done by applying Principal Components Analysis [13] with VARIMAX rotation using SPSS program [23]. The goal of this analysis is removing those indicators that bring similar information to the solution from the monitored set. Also the correlation analysis has played a great role in evaluating the indicators. The correlation coefficient value, which could be calculated for any pair of variables, lies in the range of [-1,+1] and testifies among other things, to the strength of linear relationship between the two variables. Rotated Component Matrix can be also calculated and used for explaining the relationship between the indicators and their groups. Cronbach's Alpha should be extracted for each indicator group (factor).

Finally, the overall sustainability index for ESG and economic dimension can been calculated according to the equation (2).

$$I_{UPV} = Sw_{ENVI}SI_{ENVI} + Sw_{SOC}SI_{SOC} + Sw_{CG}SI_{CG} + Sw_{EKO}SI_{EKO} \tag{2}$$

Where I_{UPV} is the overall system's index. SI_{ENVI}, Sw_{ENVI} are the sustainability index and weight of environmental dimension, respectively. SI_{SOC}, Sw_{SOC} are the sustainability index and weight of social dimension, respectively. SI_{CG}, Sw_{CG} are the sustainability index and weight of governance dimension, respectively. SI_{EKO}, Sw_{EKO} are the sustainability index and weight of economic dimension, respectively.

The sustainability index for each group of environment, social, governance and economic performance are expressed in the following equations (3).

$$SI_{Nind} = Subindex_{Nind_1} + Subindex_{Nind_2} + ... + Subindex_{Nind_n} \tag{3}$$

Where the index of each indicator is computed using the equation (4).

$$Subindex_{ind_n} = Lw.r_{ind} \tag{4}$$

Where $Subindex$ is the index of the corresponding indicator, Lw is the weight of the indicator, r_{ind} is the rate of the indicator, $Nind$ is indicator for the dimension (environment, social, governance or economy).

3 The proposed neural networks for sustainability assessment

Artificial neural networks can be implemented in the field of automated evaluation of a company's financial situation; i.e., in identifying the company's actual condition (e.g., under-capitalized, over-capitalized, non-descript situation) [20]. Upon identifying the company's condition (its financial indicators), an expert or perhaps an expert system can formulate incentives leading to a considerable limitation of entrepreneurial risks, and in the long run, can contribute to the company's stability [6]. Identifying the company's situation is a fairly weighty problem in the case of smaller companies, especially when the company does not employ a financial expert [27]. The possibility of evaluating a large number of descriptive indicators (in

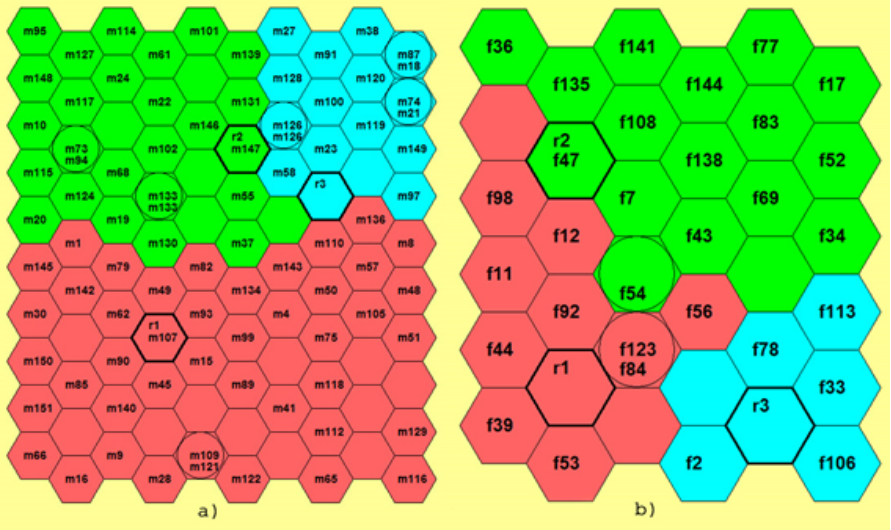


Figure 2 The result of enterprise sustainability assessment.

this case: financial indicators), which are apposite concerning the situation of the object-company [7], is one of the advantages of neural networks. In contrast with evaluations performed by people-experts, we do not have to accede to a reduction of this descriptive data (their large number, the temporal difficulty of the evaluation); with regard to this situation, we are able to achieve, in the case of unambiguously identifiable cases (under-capitalized or over-capitalized state), a considerably high speed and accuracy. With nondescript conditions, it is necessary to call for subsequent expert assessment [26].

Prior to the classification itself, it is necessary to teach the neural network[4]. Here, however, we must distinguish whether the teaching data are available–including a division of the companies into classification groups (the method of teaching with the teacher)–or not (teaching without the teacher) [22]. This situation influences the selection of the suitable neural network type and of the teaching algorithm. In the first case, if the company's history data, including their differentiation, are readily available, we can employ the multi-layer neural network and the Back Propagation Errors teaching algorithm [24]; in the opposite case, we can use self-organizing neural networks and Kohonen's teaching [19].

[4]The supposition is that we can employ data from the neural network teaching area; subsequently we make use of the principle of the universal neural-network generalization

The situation resolved by means of this type of network (a self-organizing network) is displayed on Figure 2 for the option of a smaller and greater number of classified companies. We are speaking of agricultural companies coming from Czech Republic's manufacturing industry. The graphic layout allows us to determine the objects pertinence to individual sets (the $r1$, $r2$, $r3$ representatives) and possibly it enables us to identify objects located on the classification borderline. Financial indicators subsequently used for determining the companies' liquidity, profitability, indebtedness and activity [26] have been used–in individually examined samples–for the division into groups. Upon determining the semantics of individual groups (by an expert) we can use this data also for establishing the multi-layer neural network's teaching. In this case, we are dealing with a multi-layer classification [14]. If this model is complemented with a suitably configured expert system, we may subsequently allocate a recommendation concerning the companys sustainability. The emphasis is put on simplifying and accelerating the resulting classification, as a part of the company's expert evaluation.

4 Conclusion

The main components of sustainability assessment are indicators. These indicators have influenced the company strategy and show how the company approaches a comprehensive performance evaluation, the management, the corporate reporting, sustainability, etc. Sustainability Reporting (SR) is an organization's voluntary activity with two general purposes: to assess the current state of an organization's economic, environmental and social dimensions, and communicate an organization's efforts and sustainability progress to its stakeholders. This assessment helps the organization to be more successful than its competing organizations; it also helps demonstrate the company's weak points and ways of improving on them.

The use of neural networks has been demonstrated on the issue of classifying a company's financial situation. The resulting implementation of this approach can be modified, depending on the nature of the task (the indicators of the company's financial situation, descriptive indicators). In both cases, it is mandatory to have access to historical descriptive data which are necessary for teaching the selected neural network type. In connection with an expert system, we can subsequently create recommendations influencing the further shaping of the given company's financial policies.

Acknowledgements

This paper is supported by the Czech Science Foundation. Name of the Projects: Construction of Methods for Multi-factor Assessment of Company Complex Performance in Selected Sectors. [Nr. P403/11/2085] and Measuring corporate sustainability performance in selected sectors [Nr. 14-23079S].

References

[1] Chvátalová, Z., Kocmanová, A., Dočekalová, M.: Corporate Sustainability Reporting and Measuring Corporate Performance. *Environmental Software Systems, Frameworks of eEnvironment. 9th IFIP WG International Symposium, ISESS 2011.* pp. 245-254. Springer, Heidelberg, 2011.

[2] EMAS, 2014 *ISO 26000 Social responsibility.* [online]. Available at: <http://ec.europa.eu/environment/emas/index_en.htm>.

[3] G3.1, 2011: *G3.1 Guidelines.* [online]. Available at: <https://www.globalreporting.org/reporting/sector-guidance/sector-guidance>.

[4] *G4 Guidelines: Reporting principles and standard disclosures2013.* [online]. Available at: <https://www.globalreporting.org/resourcelibrary/GRIG4-Part1-Reporting-Principles-and-Standard-Disclosures.pdf>. (last accessed 10.5.2014).

[5] *G4 Guidelines: G4 sustainable reporting Guidelines: Implementation manual 2013.* [online]. Available at: <https://www.globalreporting.org/resourcelibrary/GRIG4-Part2-Implementation-Manual.pdf>. (last accessed 10.5.2014).

[6] Grunwald, R., and Holečková, J.: *Finanční analýza a plánování podniku.* Oeconomica, Praha, 2002.

[7] Haykin, S.: *Neural networks and learning machines.* Pearson, Upper Saddle River, N.J., 2009.

[8] Hřebíček , J., Soukopová, J., Štencl, M., Trenz, O.: Integration of Economic, Environmental, Social and Corporate Governance Performance and Reporting in Enterprises. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **59** (2011), 157–177.

[9] Hřebíček , J., Soukopová, J., Štencl, M., Trenz, O.: Corporate Key Performance Indicators for Environmental Management and Reporting. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **59** (2011), 1211–8516.

[10] Hřebíček , J.,Trenz, O., Vernerová, E.: Optimal set of agri-environmental indicators for the agricultural sector of Czech Republic. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **61** (2013), 2171–181.

[11] ISO, 2014 *ISO 26000 Social responsibility.* [online]. Available at: <`http://www.iso.org/iso/iso26000`>.

[12] ISOHelpline, 2014 *Integrated Management System Documentation / Manuals for Integrated Certification.* [online]. Available at: <`http://www.isohelpline.com/ims_iso_9000_iso_14000_ohsas_18000_document_manual.htm.`>.

[13] Jolliffe, I.T,: *Principal component analysis.* Heidelberg: Springer 2002.

[14] Johnsson, M.: *Applications of Self-Organizing Maps.* InTech, USA, 2012. [online]. Available at: <`http://www.intechopen.com/books/applications-of-self-organizing-maps`>.

[15] Kocmanova, A., Hřebíček, J., Dočekalová, M., Hodinka, M.,Hornungová, J.,Chavatalová, Z.,Kubálek, T., Popelka, O., Šimberová, I., Topolová, I., Trenz, O.: *Měření výkonnosti podniku.* Brno: Littera. ISBN 978-80-85763-77-5.

[16] Kocmanová, A., Šimberová, I.: The Model of Environmental, Social and Corporate Governance Performance Indicators of a Company. *In the 17th World Multi-Conference on Systemics, Cybernetics and Informatics, Proceedings.* pp. 19-24. International Institute of Informatics and Systematics, Orland, 2013.

[17] Kocmanová, A., Šimberová, I.: Multidimensional Modeling of Social Performance Indicators in Processing Industry Companies in the Czech Republic. *International journal of math. models and methods in app. scienc.* **7** (2013), 500-507

[18] Kocmanová, A., Šimberová, I.: Modelling of CorporateGovernance Performance Indicators *Engineering Economics.* **23**(2012) pp. 485-495.

[19] Kohonen, T.: *Self-organizing maps.* Springer-Verlag, Berlin, 2001.

[20] Konečný, V., Trenz, O., and Svobodová, E.: Classification of companies with assistance of self-learning neural networks. *Agricultural Economics..* **(2)** (2010), 51-58.

[21] Křen, J.: *Metodika hodnocení trvalé udržitelnosti systému rostlinné produkce pro podmínky ČR.* Mendelova univerzita v Brně, Brno 2011.

[22] Mařík, V., Štěpánková, O., and, Lažanský, J.: *Umělá intelligence 1. díl.* Academia, Praha, 1993.

[23] Meloun, M., Militký, J., Hil, M.: *Statistická analýza vícerozměrných dat v příkladech.* Praha 2011.

[24] Rojas, R.: *Neural networks: a systematic introduction.* Springer-Verlag, New York, 1996.

[25] *SAFA: Sustainability Assessment of Food and Agriculture systems Guidelines.* Version 3.0 FAO, Rome, 2013. [online]. Available at: <`http://www.fao.org/fileadmin/templates/nr/sustainability_pathways/docs/SAFA_Guidelines_Version_3.0.pdf.`>(last accessed 10.5.2014).

[26] Trenz, O., Pokorný, M., and Redlichová, R.: Návrh modelu hodnocení podnikové finanční situace. *In Obchod a spotřebitel '07 (Stávková, J., eds.).* pp. 76-77. Konvoj, Brno, 2007.

[27] Valach, J. and Holečková, J.: *Finanční řízení podniku.* Ekopress, Praha, 1999.

Finding Optimal Variable Weights for Distance Calculation Using Metaheuristic Algorithms

Nikola Kaspříková[1]

Abstract. The notion of a distance measure is essential in many data analysis tasks. A distance is heavily used in cluster analysis, among others. There are many ways how to define the similarity (or distance respectively). One of the options for a distance definition is based on the Gower's definition of similarity. The weights for particular variables are often needed for calculation of the distance. However, in some of the situations in practice it may not be possible to get the weights directly. The problem of finding the optimal weights for particular variables for distance calculation using data on pairwise distances between items as input is addressed.

Thanks to a still increasing availability of the computation power, some computationally intensive heuristic optimization techniques have become popular. These techniques include differential evolution algorithm, its modifications and many recent biologically-inspired algorithms as well. Results of an application of a couple of general purpose heuristic methods to find optimal variable weights for distance calculation are reported and the impact of values of control parameters on the quality of the solution is discussed.

Keywords: differential evolution, distance calculation, optimization.

JEL classification: C61
AMS classification: 90C26

1 Introduction

Many problems in business practice include some optimization problems that are difficult to solve analytically (as there may be multiple local optima or there are other undesirable properties) and one has to resort to heuristic techniques. The heuristic algorithms may reach some suitable solution, even though not necessarily always the optimal one, while keeping the cost of computation at an acceptable level. Heuristic optimization methods include trajectory-based methods (there is just one candidate solution, which is modified in each iteration) or population-based methods (such methods work with a population of solutions in every iteration, which is often called a generation), as discussed e. g. by Gilli et al. in [3]. Evolutionary algorithms include the particle swarm optimization – see [5] or the differential evolution method which was introduced by Storn and Price in [8]. Recent development in the field includes many new biologically-inspired algorithms, such as the krill herd algorithm – see [2].

Heuristic methods are often computationally intensive. Such methods have recently become popular thanks to the rise of the available computation power. Values of the control parameters of these general purpose algorithms often have to be tuned for the particular problem at hand (as there does not seem to be a general rule for setting the optimal values of these parameters which would be suitable for all the objective functions and situations in practice). Regarding the differential evolution method, which is often considered to be a rather efficient tool for solving many difficult optimization tasks, the control parameters of the algorithm include the size of the population, the probability of crossover and the step size.

The notion of a distance measure is essential in many data analysis tasks in practice. A distance is heavily used in cluster analysis, among others. There are many ways how to define the similarity (or the distance respectively). One of the options for a distance definition is based on the Gower's

[1]University of Economics in Prague, Department of Mathematics,
Nám. W. Churchilla 4, 130 67 Praha 3, Czech Republic, nb33@tulipany.cz

definition of similarity (see [4] and [7]). This distance measure is quite intuitive for the users of the data analysis outputs and allows dealing with mixed variable types. The trouble is, that the weights for particular variables are often needed for calculation of the distance. However, in many situations in business practice it may not be possible to get the weights directly. On the other hand, it may be the case, that a set of pairwise distances between items considered is available. Such situation often applies e.g. in an e-commerce environment.

The problem of finding the optimal weights for particular variables for distance calculation is addressed in this paper. We seek the solution with the use of selected heuristic algorithms - the local search method, the threshold accepting method and the differential evolution method - and we are using the data on pairwise distances between items as the input for calculations. The Spearman rank correlation coefficient between the source distances and the distances in the resulting solution is used as the objective function to assess the quality of the solution and the results of the application of these general purpose heuristic methods are evaluated.

The organization of the paper is as follows: the objective function used in the analysis and the data set used are described first, then the principles of selected heuristic optimization methods are recalled in the Material and Methods section. Then the results of the analysis are reported.

2 Material and Methods

The dataset used in the experiment has 10 variables and 1000 cases. The Gower's distance (for information on distances see e.g. ([4], [9] or [7]) is calculated for each (unordered) pair of cases using a particular chosen weights vector to aggregate the information on dissimilarity between the two cases with respect to particular variables. That means that the solution in our optimization problem is going to be represented by a weights vector which has 10 components. In practice, it is sufficient to work with components which have values in range [0, 1].

2.1 The Objective Function

There exist several reasonable objective functions for evaluation of the quality of the solution and the choice of an appropriate function depends heavily on the way how the results of the analysis are supposed to be used and on the ultimate purpose of the analysis. In this experiment, we use the Spearman rank correlation coefficient between the distances calculated using the original variables weights vector (which is known in our experiment, but is often likely to be unknown in practice) and the distances calculated using the weights vector currently considered as a solution.

2.2 Optimization Methods

The threshold accepting method is an optimization heuristic method that has been introduced in [1] and it operates using the following general scheme for minimization of the objective function (here denoted OF):

1. the initial solution x_0 is generated and the current solution x_c is assigned the value x_0
2. repeat until the limit of the number of iterations is reached:
   (a) generate new (neighbour) candidate solution x_n
   (b) if $OF(x_c) + t > OF(x_n)$ set $x_c \leftarrow x_n$, otherwise keep the current x_c
   (c) return the best x_c overall

The control parameters of the threshold accepting algorithm include the (non-negative) threshold value t (which may depend on the iteration number). The function for obtaining the neighbour (i. e. the next candidate) solution has to be supplied for the particular task being solved. In this analysis a simple function was used. The solution was represented a vector which has 10 components (the initial solution was selected at random). To obtain the neighbour solution, the values of the current solution are slightly changed by adding a vector chosen at random.

The local search method may be considered as a special case of the threshold accepting method with the threshold value set to 0. The solution can not deteriorate when using the local search method, whereas when using the more general threshold accepting method, the current solution can deteriorate (i.e. the objective function value can be worse than at the previous step) to escape a local optimum.

The differential evolution method is a population-based method (several trajectories are considered at a time), and it includes several control parameters - the step size (F), the probability of the crossover (CR), the number of generations (NG) and the population size (PS). For details see the original paper on this method [8].

We use the implementation of the threshold accepting method, the local search method and the differential evolution method available in the NMOF package (see [3]) in the R computing environment [6]. Since the objective function is minimized in this implementation, as is the common approach, the correlation coefficient has always been multiplied by -1.

3 Results

The optimal 10-components weights vector for distance calculation has been searched for and the quality of the solution has been assessed using the Spearman rank correlation coefficient (which was for convenience multiplied by -1 to get the final objective function). The application of any of the selected heuristic methods (threshold accepting or local search, differential evolution) resulted in a solution with quite high value of the correlation coefficient.

The local search method running in 100 iterations has been applied first and it gave 0.9966 as the value of the correlation coefficient (see Figure 1 which shows how the objective function value changes with the iteration number).

For the threshold accepting with the following parameters: the threshold values sequence: ts = 0.02, 0.02, 0.01, 0.01 and with 25 iterations per threshold value, i. e. again with 100 iterations in total, the correlation coefficient obtained was 0.9948 (see Figure 2).

The application of the differential evolution method with the step size 0.5, the probability of crossover 0.9 and with just 20 generations and the population size 5 (to make the 100 objective function evaluations in total, similarly as when using the previous methods) resulted in the correlation coefficient value 0.9901 (see Figure 3), which again may be considered a highly satisfactory result. The objective function value was nearly the same when running the differential evolution algorithm with the number of generations and the population size set to 10. The correlation coefficient value obtained in this case was 0.9906 (see Figure 4). So it seems that the values of the control parameters did not have any substantial impact on the results in this case.

Regarding the time needed for the computations (which is generally likely to depend on the hardware and platform used), the running times in this experiment were around 70 seconds for the local search or threshold accepting method and around 80 seconds for the differential evolution method.

4 Conclusion

The results of the experiment suggest that the problem of finding the optimal weights for particular variables for distance calculation using data on pairwise distances between items as input can be conveniently solved using several metaheuristic methods. These methods allow reaching a highly satisfactory value of the objective function (which was the value of Spearman rank correlation coefficient in this case) and reach such solution quite quickly. It was shown that both the threshold accepting method (or a local search) and the differential evolution method are suitable for solving the problem.

Acknowledgements

This work was supported by the University of Economics in Prague [project number IP400040].

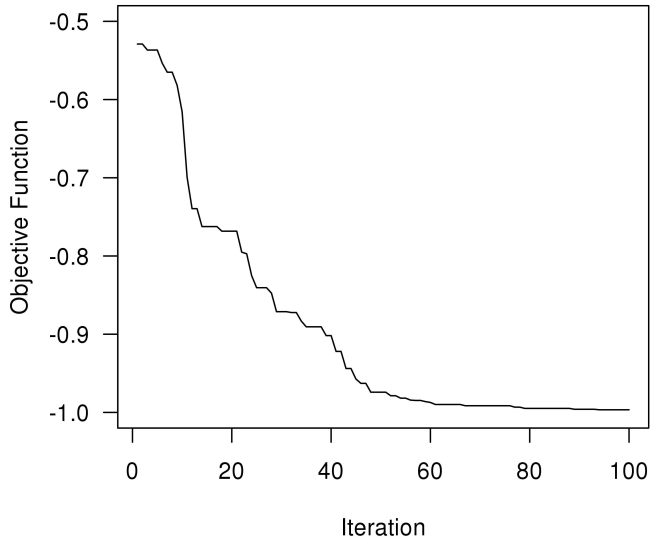


Figure 1 Local search

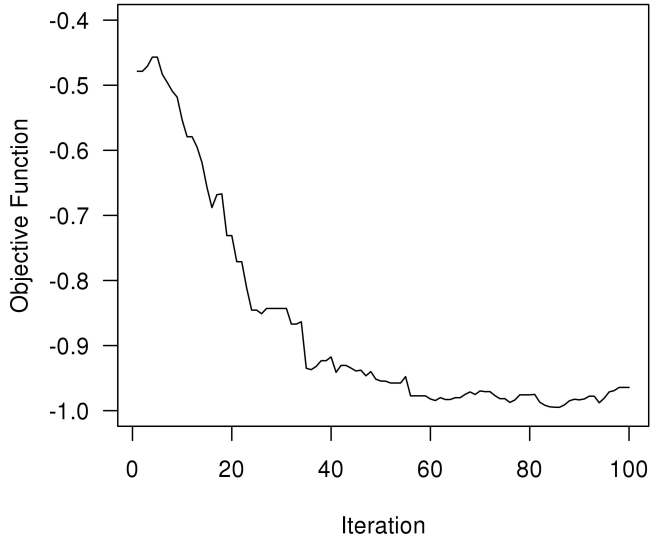


Figure 2 Threshold accepting

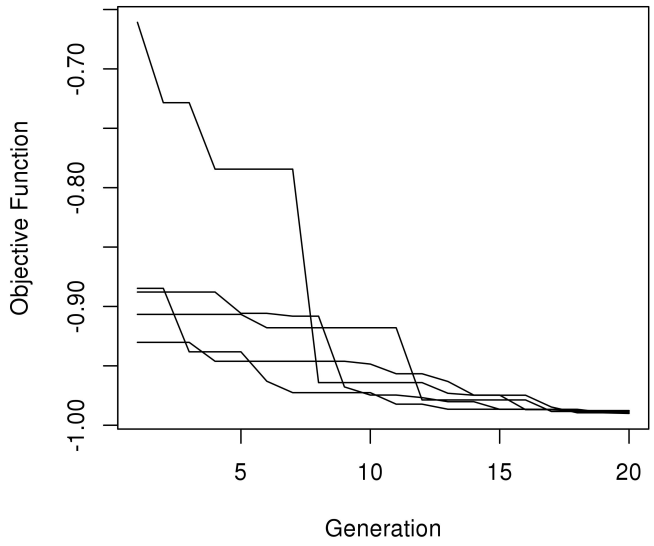


Figure 3 Differential evolution (NG=20, PS=5)

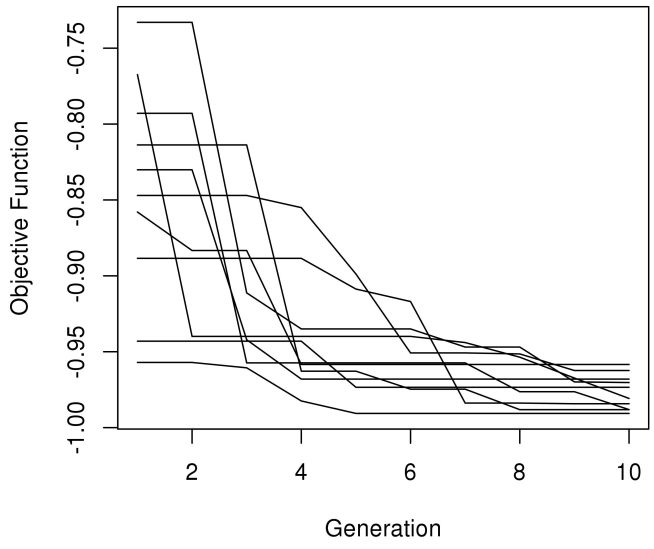


Figure 4 Differential evolution (NG=10, PS=10)

References

[1] Dueck, G., and Scheuer, T.: Threshold accepting. a general purpose optimization algorithm superior to simulated annealing. *Journal of Computational Physics*, **1** (1990).

[2] Gandomi, A., H., and Alavi, A., H.: Krill herd: A new bio-inspired optimization algorithm. *Commun Nonlinear Sci Numer Simulat*, **17** (2012).

[3] Gilli, M., Maringer, D., and Schumann, E.: *Numerical Methods and Optimization in Finance*. Academic Press, 2011.

[4] Gower, J. C.: A general coefficient of similarity and some of its properties. Biometrics **27** (1971), 857–874.

[5] Kennedy, J., and Eberhart, R.: Particle swarm optimisation In: *Proc. of the IEEE Int. Conf. on Neural Networks*, Piscataway, NJ, 1995, 1942–1948.

[6] R Core Team: *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2012.

[7] Rousseeuw, P., Struyf, A., and Hubert, M.: Clustering in an Object-Oriented Environment. *Journal of Statistical Software*, **4** (1996).

[8] Storn, R., and Price, K.: Differential Evolution A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *Journal of Global Optimization* **11** (1997).

[9] Venables, W. N., and Ripley, B. D.: *Modern Applied Statistics with S*. Springer, 2002.

Multicriteria evaluation of employees – case study

Zuzana Kiszová [1]

Abstract. The employee evaluation is one of the most important activities of the personnel department in companies. Here, some methods in multicriteria assessment, particularly comparative methods or scale methods are usually applied.
The simple pairwise comparison method belongs to the first group of methods. The decision maker says which employee of the evaluated pair is better according to the given criterion. This employee gets the point. Final ranking is then determined by the sum of points of each employee.
The scale method is the second most applied method in employee evaluation in practice. To assess the degree of fulfilment of the criteria considered the scales are used. Simple sum of points for each employee is the result of the scale method.
Both mentioned methods are rather simple to use but there is one significant disadvantage – the different importance of the criteria are not usually taken into account.
The analytic hierarchy process (AHP) is a method applicable in various situations. When compared to the above mentioned simple methods AHP is more complex to calculate the results. It enables to consider different importance of the criteria. An illustrating example is attached.

Keywords: Employee evaluation, multicriteria evaluation, pairwise comparison, scale.

JEL Classification: C44, O15
AMS Classification: 90B50, 90C29, 91B06

1 Introduction

The evaluation of employees is one of basic activities in human resources management. Some mathematical methods are usually applied when assessing employees, their performance and behavior.

Section 2 is devoted to the simple pairwise comparison method, which belongs to easily appraised ways of evaluation. Section 3 deals with scale method often used not only in employee assessment. Analytic hierarchy process as representative of sophisticated multicriteria decision making methods is characterized in section 4. These three methods are applied on a case study of employee evaluation in section 5 and some conclusions are derived in section 6.

2 Simple pairwise comparison method

The simple pairwise comparison method (SPCM) is one of the comparative methods used in employee assessment. The main principle is to declare which of the pair of compared persons is better with regard to given criterion. The winning employee gets the point, the losing one gets zero. It is necessary to compare every couple of given group. The person responsible for the evaluation may take one or more criteria into consideration [1].

It is suitable to arrange the pairwise comparisons in a table (see table 1). Employees are termed e_i, $i = 1, 2,..., n$, the score of the pairwise comparison of the i-th employee in relation to the j-th employee is $s_{ij} \in \{0, 1\}$, $i, j = 1, 2,..., n$. Note that $s_{ii} = 0$ for all i.

If only one criterion is relevant and taken into account, the final score of i-th employee is:

$$S_i = \sum_{j=1}^{n} s_{ij}, \ i = 1, 2,..., n. \tag{1}$$

Probably more than one criterion is important to evaluate the employees. Let's have m criteria. In that case every couple of persons is pair compared with respect to the k-th criterion. The score of i-th employee in relation

[1] Silesian University in Opava, School of business administration in Karviná, Univerzitní náměstí 1934/3, Karviná, Czech Republic, e-mail: kiszova@opf.slu.cz

to the *j*-th employee according to the *k*-th criterion is s_{ij}^k, $i, j = 1, 2,..., n$, $k = 1, 2,..., m$. The corresponding sum of points of every employee is:

$$S_i^k = \sum_{j=1}^{n} s_{ij}^k, \ i = 1, 2,..., n, k = 1, 2,..., m. \tag{2}$$

The final ranking is derived from the sum by descending order.

	$\boldsymbol{e_1}$	$\boldsymbol{e_2}$	$\boldsymbol{e_3}$	…	$\boldsymbol{e_n}$	**sum**
$\boldsymbol{e_1}$	0	s_{12}	s_{13}	…	s_{1n}	$\sum_{j=1}^{n} s_{1j}$
$\boldsymbol{e_2}$	s_{21}	0	s_{23}	…	s_{2n}	$\sum_{j=1}^{n} s_{2j}$
$\boldsymbol{e_3}$	s_{31}	s_{32}	0	…	s_{3n}	$\sum_{j=1}^{n} s_{3j}$
⋮	⋮	⋮	⋮	⋱	⋮	⋮
$\boldsymbol{e_n}$	s_{n1}	s_{n2}	s_{n3}	…	0	$\sum_{j=1}^{n} s_{nj}$

Table 1 Table of simple pairwise comparison method

3 Scale method

The scale method (SM) is another simple method used in employee evaluation. The employee behavior or performance is expressed by corresponding level of the scale. Two to ten levels are usual. In general, scales are composed of even or odd number of levels, which may be expressed verbally, numerically or graphically [1]. There is one disadvantage when using odd number of levels: such a scale contains one neutral level which enables to refrain from clear attitude.

When demanding the simplicity of appraisal, the numerical scale is the most suitable form. One scale may be utilized for all criteria taken into consideration or there may be different scales for different criteria. The second option is more complicated and may become confusing.

Let's have one numerical scale containing *p* levels for all *m* criteria. The evaluation of the *i*-th employee according to the *k*-th criterion is $d_i^k \in \{1, 2,..., p\}$, $i = 1, 2,..., n$, $k = 1, 2,..., m$. The final ranking of employees is based on sum of evaluations:

$$D_i^k = \sum_{k=1}^{m} d_i^k, \ i = 1, 2,..., n. \tag{3}$$

4 Analytic hierarchy process

Analytic hierarchy process (AHP) is multicriteria decision making method utilizing both above mentioned methods.

The problem is structured in hierarchy consisting of three or levels generally. The goal of the problem is situated on the top level, the level of criteria follows. Criteria represent properties of elements on the lowest level, i.e. of alternatives. The principle of hierarchy ensures that an element located at a higher level influences elements on lower level, but not vice versa [2, 3, 4].

In employee evaluation the goal is to assess the employee performance and behavior. Criteria are characteristics related to particular work position. Alternatives to be evaluated are the employees.

Every couple of elements of given level is pair compared according to higher level and a value of the fundamental nine-point scale is assigned. Number 1 means “equal importance”, number 9 represents “extreme importance”. Pairwise comparison matrix A is created (see figure 1). Note that $a_{ii} = 1$, $i = 1, 2,..., q$.

$$\begin{array}{c} \\ \text{element } x_1 \\ \text{element } x_2 \\ \vdots \\ \text{element } x_q \end{array} \begin{array}{c} \begin{array}{cccc} \text{element } x_1 & \text{element } x_2 & \dots & \text{element } x_q \end{array} \\ \begin{bmatrix} 1 & a_{12} & \cdots & a_{1q} \\ a_{21} & 1 & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{q1} & a_{q2} & \cdots & 1 \end{bmatrix} \end{array}$$

Figure 1 General pairwise comparison matrix

The entries a_{ij} , $i, j = 1, 2,..., q$, satisfy the reciprocity property:

$$a_{ji} = \frac{1}{a_{ij}}, \text{ for all } i, j. \tag{4}$$

If n employees and m criteria are considered, the evaluator has to create $(m + 1)$ pairwise comparison matrices.

The geometric mean of rows of each pairwise comparison matrix is calculated:

$$g_i = \left(\prod_{j=1}^{q} a_{ij} \right)^{\frac{1}{q}}, \; i = 1, 2,\ldots, q, \qquad \text{satisfying } \sum_{i=1}^{q} g_i = 1 \tag{5}$$

The arisen vector represents the priority (weight) vector corresponding to given matrix. If priority of i-th employee with regard to k-th criterion calculated according to (5) is assigned u_i^k and priority of k-th criterion calculated according to (5) is assigned v^k, then final ranking is based on formula:

$$w_i = \sum_{k=1}^{m} u_i^k \cdot v^k \;, \; i = 1, 2,..., n, \tag{6}$$

where w_i is global weight of i-th employee. The employees are ranked according to their global weight in descending order.

5 Case study

Let’s have 4 employees assigned A, B, C, D. These persons are evaluated according to 3 criteria: α, β, γ. Simple pairwise comparison method, scale method and analytic hierarchy process are applied to evaluate the employees.

Table 2 contains data of simple pairwise comparisons according to criteria α, β and γ. Simple pairwise comparisons of all employees with respect to all three criteria are required.

α	A	B	C	D	sum
A	0	1	1	1	3
B	0	0	1	0	1
C	0	0	0	0	0
D	0	1	1	0	2

β	A	B	C	D	sum
A	0	0	1	1	2
B	1	0	1	1	3
C	0	0	0	1	1
D	0	0	0	0	0

γ	A	B	C	D	sum
A	0	0	0	0	0
B	1	0	1	1	3
C	1	0	0	1	2
D	1	0	0	0	1

Table 2 Data for simple pairwise comparison method

The seven-level scale is utilized for scale method. Interpretation of levels and required data are depicted in table 3. No comparisons are performed, employees are evaluated independently on each other.

level	**interpretation**
7	absolutely satisfactory
6	satisfactory
5	slightly satisfactory
4	neutral
3	slightly unsatisfactory
2	unsatisfactory
1	absolutely unsatisfactory

	α	β	γ	sum
A	7	6	3	16
B	6	7	6	19
C	5	6	5	16
D	7	4	4	15

Table 3 The 7-level scale and data for scale method

The last used method is analytic hierarchy process. Pairwise comparison matrix for criteria, weights of criteria and pairwise comparison matrices for alternatives with respect to given criterion are mentioned in table 4. Not only information about employees but information about criteria is necessary.

	α	β	γ	weight
α	1	3	6	0.644
β	1/3	1	4	0.271
γ	1/6	1/4	1	0.085

$\boldsymbol{\alpha}$	*A*	*B*	*C*	*D*
A	1	4	7	2
B	1/4	1	2	1/4
C	1/7	1/2	1	1/5
D	1/2	4	5	1

$\boldsymbol{\beta}$	*A*	*B*	*C*	*D*
A	1	1/3	2	5
B	3	1	3	4
C	1/2	1/3	1	3
D	1/5	1/4	1/3	1

$\boldsymbol{\gamma}$	*A*	*B*	*C*	*D*
A	1	1/7	1/4	1/3
B	7	1	3	5
C	4	1/3	1	2
D	3	1/5	1/2	1

Table 4 Data for analytic hierarchy process and weights of criteria

Results derived from simple pairwise comparison method, scale method and analytic hierarchy process are represented in table 5.

	SPCM		**SM**		**AHP**	
	final sum	ranking	sum	ranking	global weight	ranking
A	5	**2**	16	**2**	0.401	**1**
B	7	**1**	19	**1**	0.252	**2**
C	3	**3**	16	**2**	0.106	**4**
D	3	**3**	15	**4**	0.241	**3**

Table 5 Results

Every applied method gives different ranking. The best employee according to simple pairwise comparison method is person *B*, employee *A* follows and the shared third position belongs to employees *C* and *D*.

The first position based on scale method is identical as in the case of SPCM – the best employee of the group is *B*. The second position is shared, employees *A* and *C* seem to be equally performing according to this method. Person *D* is on the last place.

Different ranking is brought by analytic hierarchy process. The main reason is inequality of criteria significance taken into consideration. The main priority belongs to criterion α, criterion γ is minimally important (see table 4). With respect to AHP employee *A* gets the best evaluation, *B* is following. The third position belongs to employee *D* and *C* is the last one. When concentrating on global weights of employees, the minuscule difference between employees *B* and *D* is evident.

6 Conclusion

Some advantages and weaknesses of above mentioned methods emerge when applying them on example of employee assessment.

The simple pairwise comparison method is very easy to utilize and does not require special mathematical skills. The decision, which of the pair of compared persons gets the point, seems to be not very difficult but may be very subjective. The following appraisal of data may be fast. Different significance of criteria is not taken into consideration – that is the main disadvantage and the result may be distorted. The diverse extent of pairwise comparisons is not distinguished in this method.

The scale method requires prudent judgment of each employee performance. Differences between individual employees' performance are irrelevant, the satisfaction degree of requirements is evaluated. One employee may be more performing than another one, but they both may achieve absolutely satisfactory results. Similarly to the first mentioned method, the scale method does not distinguish different importance of criteria.

The analytic hierarchy process is more sophisticated than previous methods. AHP requires more calculations, but these are not difficult. It enables to take distinct priorities of criteria into account. The different extent of fulfilment of the criteria is contemplated thank to the nine-point scale used in pairwise comparisons. Globally, this method enables to express considerable extent of differences between employees and may be more rational. This is the main reason of contradictory results compared to simple pairwise comparison method and scale method. The result of AHP may be basis of consequent rewarding – the global weights of employees may represent the proportion of bonus.

The evaluator is responsible for impartiality and thoroughness of the assessment. Author of this paper thinks that sophisticated methods are more appropriate than the simple ones when evaluating employees, their behavior and performance.

Acknowledgements

This paper has been supported by GACR project No. 14-02424S.

References:

[1] Dvořáková, Z.: *Management lidských zdrojů*. C. H. Beck, Praha, 2007.
[2] Nevima, J. and Kiszová Z.: Competitiveness of Czech and Slovak NUTS 2 regions measured by AHP and equal importance method. In: *Proceedings of the 31st International Conference Mathematical Methods in Economics 2013*. VŠPJ, Jihlava, 2013, 667-672.
[3] Ramík, J. and Perzina, R.: *Moderní metody hodnocení a rozhodování*. SU OPF, Karviná, 2008.
[4] Saaty, T. L.: *The analytic hierarchy process: Planning, priority setting, resource allocation*. McGraw Hill, New York, 1994.

Index Clause Valuation under Stochastic Inflation and Interest Rate

Kamil Kladívko[1], Pavel Zimmermann[2]

Abstract. This work is focused on valuation of the reinsurer's share on a particular accident, stemming from the excess of loss (XL) reinsurance contract that applies to a general insurance annuity compensation, e.g., motor liability or workmen's compensation. The excess of loss (XL) reinsurance contract is an insurance contract between the insurer and reinsurer, which guarantees recovery payment to the insurance company for each accident in the amount of the accident that the insurance company pays off in excess of a contracted priority.

Special focus is set on the impact of the so called index clause which is usually included to the reinsurance contract. The index clause allows the reinsurer to increase the originally agreed priority by a coefficient which is, roughly said, calculated as a ratio of the sum of all nominal payments to the sum of all deflated payments.

Analytic valuation formula does not exist for the reinsurer's share without the index clause. Hence simulation model is used to perform the study. We assume a geometric Brownian motion for the inflation index and an Ornstein-Uhlenbeck process for the discount rate, where we allow the two processes to be correlated.

Keywords: Annuities, Index clause, Reinsurance, Excess of loss.

JEL classification: C44
AMS classification: 60G99

1 Introduction

This work is focused on valuation of the reinsurer's share stemming from the excess of loss reinsurance contract. Special focus is set on the impact of a specific clause, the so called index clause on the value of the reinsurer's share. We will limit our self to one accident, which is already reported. We will first specify the type of (direct) insurance products of our concern and consequently define the reinsurer's share considered.

The above mentioned reinsurance contract applies to various *insurance products*, which are marketed by insurance companies to clients. The insurance product considered is a general insurance product that guarantees regularly inflated payment until death or a specified age of the victim (or other beneficiary). Typical example of such a product is the motor third party liability in the central and eastern Europe, where certain amount of accidents are paid out regularly in the form of an *annuity* (e.g. the compensation for loss of income or care costs in the case of large bodily accidents). Another example is the workmen's compensation, commonly existing in the western Europe. Note that "business wise" (not accounting wise) these liabilities belong to non-life insurance lines, despite the fact that the amount paid off depends on survival or death of the beneficiary. This fact implies certain complications to valuation (or modeling in general) of these liabilities:

1. These liabilities are in comparison to other non-life liabilities extremely long-termed, and hence much more sensitive to discounting or inflation.
2. These liabilities are typically reinsured with non-proportional treaty (usually the excess of loss

[1]University of Economics, Prague W. Churchill Sq. 4,130 67 Prague 3, kladivko@gmail.com
[2]University of Economics, Prague W. Churchill Sq. 4,130 67 Prague 3, zimmerp@vse.cz

treaty), which is rarely encountered in life insurance annuities and the reinsurer's share may be more significant than in life insurance liabilities.

The *reinsurance contract* is an insurance contract between the insurer and reinsurer. In this contract the reinsurer agrees in exchange for financial compensation (reinsurance premium) to share the insurer's liabilities to some extend in a form of recovery payments to the insurer. In this paper, we deal with the *excess of loss* contract (XL contract). The XL contract guarantees recovery payment to the insurance company for each accident in the amount of the accident that the insurance company pays off in excess of a contracted amount. The recovery is due at the accident settlement date. The contracted amount is referred to as the *priority*.

In case of excess of loss reinsurance contract, reinsurers often do not accept the inflation risk embedded in such liabilities, and therefore include the so called *index clause* in the reinsurance treaty. The purpose of the index clause is to adjust the priority for movements of an inflation index. The index clause allows the reinsurer to increase the originally agreed priority by a coefficient which is, roughly said, calculated as a ratio of the sum of all nominal payments to the sum of all deflated payments. Notice that the reinsurance share is calculated at the time of the claim closure and therefore past inflation as well as all payments are known. Given the length of the annuity liabilities the index clause may have significant effect on the reinsurer's share and, therefore it is particularly important to include this clause in the valuation model.

The existing literature on the index clause is quite limited. Several notes on the impact of the inflation on the nonproportional reinsurance as well as reasoning for the index clause and reinsurance pricing formulas can be found in [3]. A pricing model for an excess of loss treaty including (among other clauses) also the stability (index) clause was published in [4]. An indexing method for annual aggregate deductibles and limits was scrutinized in [5]. A method for the estimate of the insurance liabilities based on individual loss level with inclusion of the excess of loss reinsurance with the index clause was developed in [2]. Deterministic evaluation of the impact of the index clause together with its analytical properties and capitalization strategy was published in [6].

The main focus of this paper is to derive the the difference in the expected present value of the reinsurer's share on technical reserve of a reported accident using a *stochastic model*. Note that as we assume that the claim is already reported, we may also assume that we already know the age and gender of the victim as well as the intensity of the compensation.

2 Reinsurer's Share and Underlying Processes

2.1 Compensation, Annuities and the Underlying Economy

We assume that the insurance product guarantees a compensation paid out continuously with a constant intensity a, to a victim with some residual lifetime. In this subsection, only fixed residual lifetime t will be considered. The following payment streams are considered:

- $A^*(0,t) = \int_0^t a du = at$: The total payment from time 0 to time t (in a constant price level).
- $A(0,t) = \int_0^t aI(u)du$: The total payment, which is continuously adjusted by an inflation index $I(u)$, specified below, from time 0 to time t. The inflation adjusted payment $A(0,t)$ corresponds to the aggregated amount paid by insurance company to the victim.

To calculate the present value of the considered cash flows, we introduce a stochastic discount rate. Specifically, the discount rate follows an Ornstein-Uhlenbeck process

$$dr(u) = \kappa\big(\mu - r(u)\big)du + \sigma dW_1(u), \quad r(0) = r_0, \tag{1}$$

where μ an unconditional mean of the discount rate, κ determines a speed at which the short-rate reverts to its unconditional mean, and σ is the volatility of the discount rate. The inflation is represented by a consumer price index (CPI), which follows a generalized geometric Brownian motion

$$dI(u) = \pi(u)I(u)du + \gamma I(u)\big(\varrho dW_1(u) + \sqrt{1-\varrho^2}dW_2(u)\big), \quad I(0) = 1, \tag{2}$$

where $\pi(u)$ may be thought of as the time dependent expected inflation rate, and where γ is the volatility of the CPI. The CPI is exposed to shocks that drive the discount rate, and in addition to shocks that are orthogonal to the discount rate, whereas ϱ determines the correlation between the CPI and the discount rate.

To sum up, our naive economy is generated under the real-world probability measure by the two independent Brownian motions $W_1(u)$ and $W_2(u)$. In particular, we do not assume an existence of a market with financial instruments, such as nominal and real bonds, which could be used for hedging the considered insurance and reinsurance products.

2.2 Reinsurer's Share

The the two reinsurance products will now be defined. In what follows, we repeatedly use the following manipulation of the max function: for any $x \geq 0$, $(x(y-z))^+ = x(y-z)^+$.

1. *Reinsurer's share with an index clause.* At a given time t, the reinsurer's payoff is given as

$$C(t) = \left(A(0,t) - \frac{A(0,t)}{A^*(0,t)}K\right)^+ = \big(at - K\big)^+ \frac{1}{t}\int_0^t I(u)du, \tag{3}$$

   where K is the priority of the reinsurance contract, and the fraction $A(0,t)/A^*(0,t) = (1/t)\int_0^t I(u)du$ represents the index clause, which removes the inflation risk from the priority. The max part of the payoff is deterministic and thus known at time 0. The only stochastic part of the payoff is time average of the inflation process. This time average can be interpreted as an Asian forward contract on the inflation process with delivery price set to 0, or alternatively as a zero-strike arithmetic Asian call on the CPI.

2. *Reinsurer's share without an index clause.* At a given time t, the reinsurer's payoff is given as

$$W(t) = \big(A(0,t) - K\big)^+ = at\left(\frac{1}{t}\int_0^t I(u)du - \frac{K}{at}\right)^+, \tag{4}$$

   which can be interpreted as an arithmetic Asian call option on the CPI with a fixed strike equal to $K/(at)$.

2.3 Mortality Risk

To this point we have considered the reinsurer's share for a fixed residual lifetime t. We now introduce a random variable T_x, which measures a random time at which an x-year old victim dies (residual lifetime). We follow the standard assumption of independence between the mortality and economic variables, in particular the probability distribution of T_x is independent of the Brownian motions $W_1(u)$ and $W_2(u)$.

The Makeham's model is assumed for the survival function of lifetime of the population of newborns, ${}_tp_0 = P(T_0 > t)$. Namely the following survival function is assumed:

$${}_tp_0 = bs^t g^{c^t}, \tag{5}$$

where b, s, g and c are parameters. The corresponding probability density function is denoted $f_x(t)$. The (conditional) density of the residual lifetime of an individual being already x years old is then

$$f_x(t) = \frac{f_0(x+t)}{{}_xp_0} = -\frac{d}{dt}\frac{{}_{x+t}p_0}{{}_xp_0}. \tag{6}$$

2.4 Valuation of the Reinsurer's Share

The standard approach to calculating the reinsurer's share value is based on taking the expectation of the discounted reinsurer's share payoff. Therefore, based on (3) and (4) we get for a fixed nonrandom

time t the respective valuation formulas:

$$\text{With IC: } C(0,t,I_0,r_0,K) = \big(at-K\big)^+ \mathbb{E}\left[e^{-\int_0^t r(u)du}\frac{1}{t}\int_0^t I(u)du\right]. \tag{7}$$

$$\text{Without IC: } W(0,t,I_0,r_0,K) = at\mathbb{E}\left[e^{-\int_0^t r(u)du}\left(\frac{1}{t}\int_0^t I(u)du - \frac{K}{at}\right)^+\right]. \tag{8}$$

Note that we calculate the reinsurer's share value at time 0, i.e., we assume that the accident has just occurred. It is straightforward to show that the value of the reinsurer's share at some later time $\tau > 0$ can be expressed in terms of a newly reported accident of the same type, but with an altered priority and intensity of the payout.

Further note that the expectation operator $\mathbb{E}[\cdot]$ in the valuation formulas is taken with respect to the Brownian motions $W_1(u)$ and $W_2(u)$. Since the two Brownian motions are assumed to be independent of the mortality risk, it is straightforward to incorporate the random time T_x into the valuation. Specifically, given our mortality model, the respective valuation formulas incorporating the uncertain maturity of the reinsurer's share are as follow:

$$\text{With IC: } C(0,T_x,I_0,r_0,K) = \int_0^\infty C(0,t,I_0,r_0,K)f_x(t)dt. \tag{9}$$

$$\text{Without IC: } W(0,T_x,I_0,r_0,K) = \int_0^\infty W(0,t,I_0,r_0,K)f_x(t)dt. \tag{10}$$

Despite an enormous effort, which has produced many academic papers, a search for an analytical solution for the price of an arithmetic Asian call has not been successful yet, not even for a deterministic discounting. Therefore, we use Monte Carlo methods to evaluate the expected values of (7) and (8). However, we would like to note that we have obtained an almost analytical solution (modulo fast and accurate numerical integration) for the reinsurer's share with an index clause under the here considered setting. The solution will be presented in a subsequent paper, where also an economy with a complete financial market is introduced, and therefore a risk-neutral valuation approach to the reinsurer's share is developed.

3 Parameter Calibration

As already mentioned above, we evaluate the reinsurer's share under the real-world measure. We therefore estimate parameters of the discount rate and CPI processes (1) and (2), respectively, from the available time series data. The CPI is published by the Czech Statistical Office on a monthly basis. We adjust the CPI for seasonal effects to measure its correlation with the three-month PRIBOR, which is used as a proxy for the discount rate. For simplicity, we assume that the expected inflation rate $\pi(u)$ is a constant π. The estimated parameters, reported in the left-hand part of Table 1, are found by maximum likelihood based on monthly data from the beginning of 1998 until April 2014. The estimated parameter values are relatively stable for different sampling periods. However, the maximum likelihood estimation of the unconditional mean, μ, of the discount rate results is in positive values very close to zero, or even in negative values. This is due to an apparent downward trend in the three-month PRIBOR time series during the considered period. Since we believe that interest rates must be mean reverting, we set μ equal to 0.04.

Generation life tables constructed in [1] updated for the base values of 2007 are used. Parameters of the Makeham's function are fitted using ordinary least squares approach to the table number of survivals up to age x, denoted as l_x. (Technically, we are not fitting the survival probability but the survival function multiplied by the fixed table radix $l_0 = 100000$.) The survival probabilities are then calculated based on the formula ${}_np_0 = l_n/l_0$ and for the discrete time intervals dt, the probability of death in an interval $[t, t+dt]$ is:

$$P(t < T_x < t+dt) = \frac{{}_{x+t}p_0 - {}_{x+t+dt}\,p_0}{{}_xp_0} = \frac{l_{x+t} - l_{x+t+dt}}{l_x}. \tag{11}$$

The estimated values of the parameters are displayed in Table 1.

κ	μ	σ	π	γ	ϱ	b	s	g	c
0.30	0.04*)	0.006	0.024	0.010	0.19	100040	0.99878	0.99997	1.12310

Table 1 The left-hand part reports the estimated parameters of the underlying economic processes. *) The unconditional mean of the discount rate has been set to 0.04 rather then estimated. The right-hand part reports the estimated parameters of the Makeham's model fitted to l_x.

4 Numerical Results

The results are organized as follows: We start with the results for a "base case", i.e., with parameters displayed in Tables 1 and 2, which are considered as a "typical example". Then we perform particular single parameter stresses to illustrate the sensitivity of the reinsurer's share on these values. All parameters, except for ϱ, are shifted up and down by 50 %. In all cases time step $dt = 1/12$ is used and 10 000 simulations were performed. Further on the parameters from Table 2 were set as 'the base case'.

a	K	x	Sex
0.5	20	30	Male

Table 2 Base case parameters. (a and K in mil. CZK.)

The results for the base case are displayed in the first row of Table 3. It is obvious that for the base case the index clause has substantial impact on the reinsurer's share, and hence on the net reserve. It is therefore essential to take this clause in consideration when evaluating reinsurance treaty or adequacy of reserves. The second line of this table displays the results when only deterministic model is applied, i.e., when both volatility parameters γ and σ are set to 0. The impact of the stochastic parameters is apparently quite low both on the values of the reinsurer's share as well as on the impact of the index clause (IC).

The first stress test is performed on the payout intensity a. It is obvious that for the down shift ($a = 250000$), the index clause annihilates the reinsurer's share. As the payout intensity increases, the reinsurer's share increases and the relative impact of the IC is decreasing. So for the highest accidents, the impact of the IC is partially mitigated.

The second stress test is performed on the inflation drift π. Although this parameter is of course significant for the value of the reinsurer's share itself, the impact on the difference between contract with and without IC is not so dramatic. For the "up shift" ($\pi = 0.036$), the increase of the relative impact of the IC is only around 6 percentage points and when π is further increased, the ratio is not increasing substantially.

The third stress test is performed on the long term mean of the nominal risk free rate μ. As the cash flows are quite long, discounting has a substantial impact on the reinsurer's share itself. But the impact on the difference between the contract with and without the IC is quite negligible.

The fourth stress test is performed on the correlation between the inflation and nominal risk free rate ϱ. Here the down shift is set to 0 (independence assumption) and the up shift is set to 0.5. It is quite surprising, that this parameter has only a very little impact on the reinsurer's share as well as the difference between the contracts with and without the IC. It is in fact quite important because the dependence assumption complicates theoretical analysis of the modeled phenomena substantially and approximation with the independent case may be a helpful benchmark.

5 Conclusions

The index clause is usually present in the excess of loss reinsurance contracts. Based on the above mentioned results, it is obvious that on reasonably large accidents, the impact of this clause on the reinsurer's share (and hence the net reserve) may be very substantial. On the other hand, stress tests indicate that although the sensitivity of the value of the reinsurer's share it self may be significant, the impact of the index clause is generally quite robust.

Stressed	Scenario	Reinsurer's share		Relative impact on the base case		Impact of IC	
		With IC	Without IC	RS with IC	RS Without IC	CZK	Relative
-	Base case	1.45	3.61	-	-	2.16	59.8%
-	Deteministic	1.44	3.58	-0.9%	-0.7%	2.14	59.9%
a	0.75	4.32	6.62	198.4%	83.4%	2.29	34.7%
[0.5]	0.25	0.00	0.80	-100.0%	-77.7%	0.80	100.0%
π	0.036	2.29	6.67	58.3%	85.0%	4.38	65.6%
[0.024]	0.012	0.96	1.77	-33.9%	-50.9%	0.81	45.9%
μ	0.06	0.48	1.27	-66.8%	-64.8%	0.79	62.2%
[0.04]	0.02	4.44	10.53	206.7%	191.8%	6.08	57.8%
ϱ	0	1.45	3.62	0.4%	0.4%	2.17	59.9%
[0.19]	0.5	1.45	3.61	0.2%	0.1%	2.16	59.8%

Table 3 Reinsurer's share and the influence of the IC for the base case and relevant ± 50 % parameter shifts. The base case value of the parameter is in [ ]. All monetary values are in mio CZK.

Additionally the sensitivity on the volatility parameters is relatively low and the difference between stochastic and deterministic approach is quite negligible. The correlation of the error term of the nominal risk free rate and the inflation index has also very little impact on the results. These results then suggest that deterministic approach would also be an acceptable alternative for valuation.

Acknowledgements

Supported by the grant No. P404/12/0883 and P402/12/G097 of the Czech Science Foundation.

References

[1] Cipra, T.: Cohort life tables for pension insurance and pension funds in Czech Republic (in Czech), *Pojistne rozpravy* **3** (1998), 31–57.

[2] Drieskens, D. et al.: Stochastic projection for large individual losses. *Scandinavian Actuarial Journal* (2010), 1–39

[3] Ferguson, R. E.: Nonproportional Reinsurance and the Index Clause. In: *CAS Proceedings* **LXI** (1974), 141-164

[4] Walhin, J. F. Herfurth, L., De Longueville, P.: The practical pricing of Excess of Loss treaties: actuarial, financial, economic and commercial aspects. *Belgian Actuarial Bulletin* **1** (2001), 40–57.

[5] Yeung, K. C. (2011) Index Clause for Aggregate Deductibles and Limits in Non-Proportional Reinsurance. In: *Casualty Actuarial Society E-Forum, Spring 2011*, (2011), 1

[6] Zimmermann, P.: Index clause: analytical properties and the capitalization strategy, *European Actuarial Journal* **2** (2012), 149–160.

Is it possible to use multi-criteria criteria decision making methods for rural typology?

Renata Klufová[1], Michael Rost[2]

Abstract. The period 2007-2013 was characterized by an emphasis on a common strategy of rural development for the whole EU. In spite of common European and national rural development strategies there is still scope for individual strategies of municipalities and rural areas in terms of their own way of development. There are many approaches to defining rural areas - classifications of rural areas cannot be made without the awareness of the dichotomy of urban-rural areas, which is perceived by many authors rather as a kind of continuum. Various methods of multi-criteria statistical analysis are usually used for rural typology. The paper focuses to demonstrate the usage of multi-criteria decision making methods for this typology. Weighted Sum Approach is applied to chosen socio-economic criteria in order to make Czech rural typology.

Keywords: Multicriteria Decision Making, Weighted Sum Approach, Cluster Analysis, Rural Typology, GIS

JEL classification: C44
AMS classification: 90B50

1 Introduction

Rural areas occupy from 73% - 82% of the area of the Czech Republic (depending on specific criteria) and about 22% - 29% of the total population live there. Common characteristics for those areas are lower level of economic activities and lower density. Significant regional differences have become evident in the Czech rural space. They are primarily conditioned by geographical location, socio-economic and natural conditions. They are also manifested by different economic productivity and interest or disinterest of their inhabitants. In order to experience various rural problems, it is necessary to be concerned with differences of social, economic, environmental, cultural and historical aspects. Solving of known problems needs a complex approach with respect to wider territorial and socio-economic relations, which support trends of sustainability [1].

Rural space becomes differentiated in view of the fact that a relative geographic location changes, the settlement system differentiation and hierarchy is deepening [5]. Differences can be seen in a population structure as well as in rural activities. It leads to different development preconditions of rural villages. The typology discussed in this article works on the assumption that a development potential projects into chosen socio-economic characteristics. Available statistical data has been used for the analysis (rural population structure at the level of municipalities from census 2011, employment, geographic location etc.). Variables describing quality of social and human potential were also put into the analysis.

We follow new approaches in regional development here. These approaches come out from the institutional theories and understand a possibility of local actors support as a development precondition. They also emphasize a necessity of the usage of local conditions and development assumptions as well as endogenous potential activation. Specification of distinctive rural types should than serve not only for a deeper recognition of a current state or its determining processes, but especially for a cognition of a further development preconditions. Consequently it could be used for the design of different tools of rural development.

[1]University of South Bohemia/Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, klufova@ef.jcu.cz

[2]University of South Bohemia/Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, rost@ef.jcu.cz

1.1 Defining and typologing rural areas

There are many approaches to defining rural areas - classifications of rural areas cannot be made without the awareness of the dichotomy of urban-rural areas, which is perceived by many authors rather as a kind of continuum (see e.g. [2]). Some authors have studied the influence of the large urban areas on their background. In this context, the term "urban shadow" was used (e.g. [4, 6, 7]). It is used to express the various ways in which the effect of urbanization was spreading further beyond the borders of built-up areas which represents the important dimension of any typological classification of the rural areas.

The interest has recently shifted from material aspects of rural areas to the political-economic structures and social construction. This requires new research questions, methodology and opportunities for new interdisciplinary connections. The attention is focused on the features of the rural identity, exclusion and inclusion, and the commodification of rural tourism experience, which were often put aside.

In the Czech Republic there is often used the typology created by Perlín [12]. It is based on a specific historical development and current socio-economic indicators and divides the countryside at the district level. The author himself [13], however, mentions the problematic nature of such created typology, caused by the limited data base. Therefore we should not ignore the currently created typology [13], which has been created within the project MMRWD-01-07 entitled the Regional differentiation of rural municipalities in the Czech Republic: disparities and development opportunities. By using methods of multivariate statistics (principal components, cluster analysis and many others) there were, at the level of municipalities with authorized municipal office, identified eight types of rural areas in the Czech Republic: core countryside - in the hinterland of the largest cities and along major development axes, Moravian peripheries a unique type located in the historical lands of Moravia and Silesia, recreationally problematic countryside large parts of Czechias borderlands that were subject to resettlement after WWII, intensive recreational areas located in mountainous parts of the borderlands, such as Krkonoše, Šumava or Beskydy, structurally affected industrial countryside found especially in structurally affected regions of northern Bohemia, non-specified countryside.

Klufová has also carried out a typology of the Czech rural space based on newer data from census 2011 [10]. She used factor and cluster analysis which have led to six types of the Czech countryside shown in the picture 1 where is shown typology of rural municipalities with the number of inhabitants in 2011 above 3 000 people.

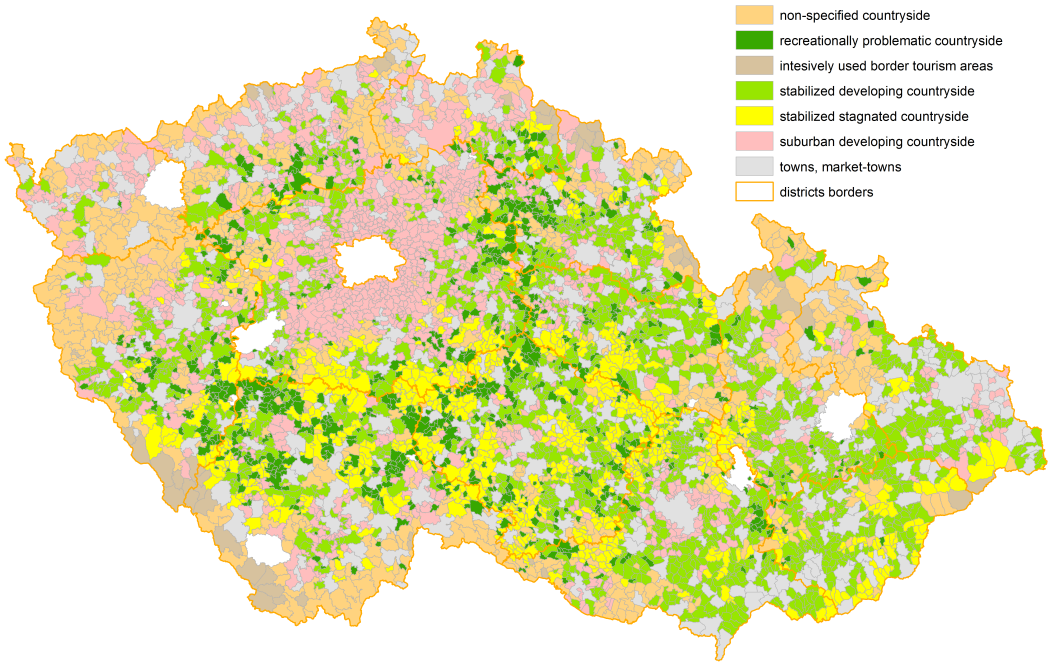


Figure 1 Typology of the Czech rural municipalities based on factor and cluster analysis

2 Material and Methods

When creating rural typology we came out from basic characteristics and questions connected to the rural development: Size and geographic position; agricultural production - how is a proper occupation for inhabitants which don't work in agriculture or forestry?; Space for recreation - who will live in the countryside and who will use the it? Does tourism ensure jobs for rural inhabitants? Space for living - cheaper housing, jobs in towns; Where will inhabitants live - value of social and community life; Space for life - do community and technical facilities correspond to this function? - job opportunity offer, enterprise promotion; quality environment, nature.

Characteristics and questions mentioned above were expressed by a set of variables available at the level of municipalities:

- size and geographic location - number of inhabitants in 2011 (O2011), municipality cadastral area (HA), population density 2011 (hustota11), number of municipality's parts (CASTI11), geographic location[1] - distance from the nearest municipality with extended power (NEARORP), distance from the nearest railway station (NEARZS), distance from the nearest main road (NEARHSIL);
- employment - employment in primary sector (PRIM11r), the share of empolyees of the total number of employed (ZAMST11r), share of persons working on own account (OSVC11r), share of agricultural land (ZP11r), rate of employment (OZC11r - share of employed persons of economically active population, ZMC11r - share of employers of the total number of employed);
- space for recreation - employment in the branch of accommodation, catering and hospitality (UBYT11r), number of collective accommodation establishments (HUZ11), share of unoccupied houses of the total number of houses (NEOBD11r), share of unoccupied houses used for recreation (DREKR11r), share of water and forest areas of the total cadastral area (VODYLESY11r);
- space for housing - share of built-up and other areas(ZAST11r), share of family houses (RD11r), share of houses built-up or reconstructed in the period 2001–2011 (D01_11r), share of houses atteched to sewerage (KANAL11r), share of native inhabitants (NAROB11r), net migration - chronological average 2001–2012 (chpMS01_12), natural increase - chronological average 2001–2012 (chpPP01_12), index of number inhabitants development 2001–2011 (IO01_11), share of married persons (MANZ11r), share of believers in God (VIRA11r), index of ageing (IS11), index of economic burden (IEZ11)
- space for life - share of daily commuting persons (VYJdmo11r), index of education (IVZD11).

The own typology has been made in the subset of municipalities. Military training areas and Prague have been excluded from the analysis. Rural municipalities cannot be evaluated without information about their position in the whole set of municipalities. Thus we assessed the regional potential for all of them with the exception of military training areas and Prague as outliers. The first step was a correlation analysis. We have found that low or intermediate correlations prevail among the variables. Three variables ZP11r, chpPP01_12 and OSVC11r were than excluded due their high correlations because they seem to be redundant.

Cluster analysis of the chosen variables has been implemented in order to identify clusters of interrelated variables. Hierachical clustering by Ward methods (see e.g. [9]) based on Pearson correlation coefficient identified 4 clusters which describe certain fields of development potential. They can be described in a following way:

1. municipality size and importance - hustota11, ZAST11r, KANAL11r, HA, O2011, CASTI11, HUZ11, UBYT11r;
2. municipality development - IVZD11, BDINT11r, OZC11r, ZMC11, D01_11, chpMS01_11, IO01_11;
3. potential for economic and leasure activities - IS11, IEZ11, PRIM11, NEOBD11r, DREKR11r, NEARHSIL, NEARORP, NEARZS, , VODYLESY11r;
4. social stability and employment - ZAMST11r, NAROB11r, VIRA11r, VYJdmo11r, RD11r, MANZ11r.

[1]Distances from the neares municipality with extended power, railroad station and main road (highway, limited-access highway or main road) have been calculated by NEAR tool in ArcGIS 10.1.

Various rural typologies, not only in the Czech republic but also for example in the EU, are usually based on big set of variables and their analysis by multivariate statistical methods. We decided to assess a position of Czech municipalities according their development potential in another way. The main question answered could be stated as: "Is it possible to obtain realistic typology by methods of multi-criteria decision (MCDM)? Can we obtain similar results as by the use of multivariate statistical methods?". We decided to use the Weighted Sum Average method (see for example [3]) which is based on linear utility function:

$$u_i(x) = \sum_{j=1}^{n} w_j r_{ij}(x), \tag{1}$$

where $u_i(x)$ denotes the utility of the ith alternative (municipality) and $i = 1, 2, \ldots, n$, w_j denotes the weights of the jth criterion; $r_{ij}(x)$ is the normalized preferred rating of the ith alternative (municipality) with respect to the jth criterion for all commensurable units. All criteria are assumed to be independent. In addition, the normalized preferred ratings $r_{ij}(x)$ are defined by:

$$r_{ij}(x) = \frac{x_{ij} - \max\limits_{j} x_{ij}}{\max\limits_{j} x_{ij} - \min\limits_{j} x_{ij}}. \tag{2}$$

This method computes the global utility of the alternatives as the weighted sum of normalised criterion values. The method provides complete ranking of alternatives according to their global utilities [8, 11, 14].

Four utility functions were calculated for each cluster of variables. The weights of the particular variables in clusters have been determined according to their importance in clustering mentioned above (based on Pearson correlation). Statistica SW was used for cluster analysis, utility functions were calculated in Excel. Cartograms (figures 1,2) were created in ArcGIS SW.

Thus we obtained four utility functions describing four fields of the development potential mentioned above. Values of all these functions served as an input to the cluster analysis of municipalities which led to a new typology according the development potential. We used k-means methods based on squared Euclidean metrics to obtain seven different types of rural municipalities.

3 Results and Discussion

Six identified clusters have been named according to their particular characteristics (see table 1 and picture 2):

1. **stabilized developing countryside** - population-larger municipalities with high population density and relatively favorable age structure. These municipalities predominantly lie in Moravian part of the Czech Republic, in Vysočina district.

2. **suburban countryside** - this consists of municipalities located in the hinterland of the largest cities and along major development axes and main transport corridors that connect central areas,

3. **structurally affected industrial countryside** - specific type of rural municipalities to be found especially in structurally affected regions of northern Bohemia,

4. **intensively used border tourism areas** - significant dominance of recreation and related activities, located in mountainous parts of the borderlands, such as Krkonoše, Šumava or Beskydy,

5. **stabilized stagnating countryside** - this type includes a large part of the Bohemian interior, especially south of Prague and to the East, until the Českomoravsk´ vrchovina [Bohemian-Moravian Highlands]. It consists mainly of economically weak areas, located near the borders of major administrative regions (the so-called inner peripheries).

6. **non-specified countryside** - this final type consists of rural areas that are not included in any of the previous types.

Cluster 1 labeled as "stabilized developing countryside" reached average values of the utility functions "municipality size and importance, development and potential for economic and leisure activities". The

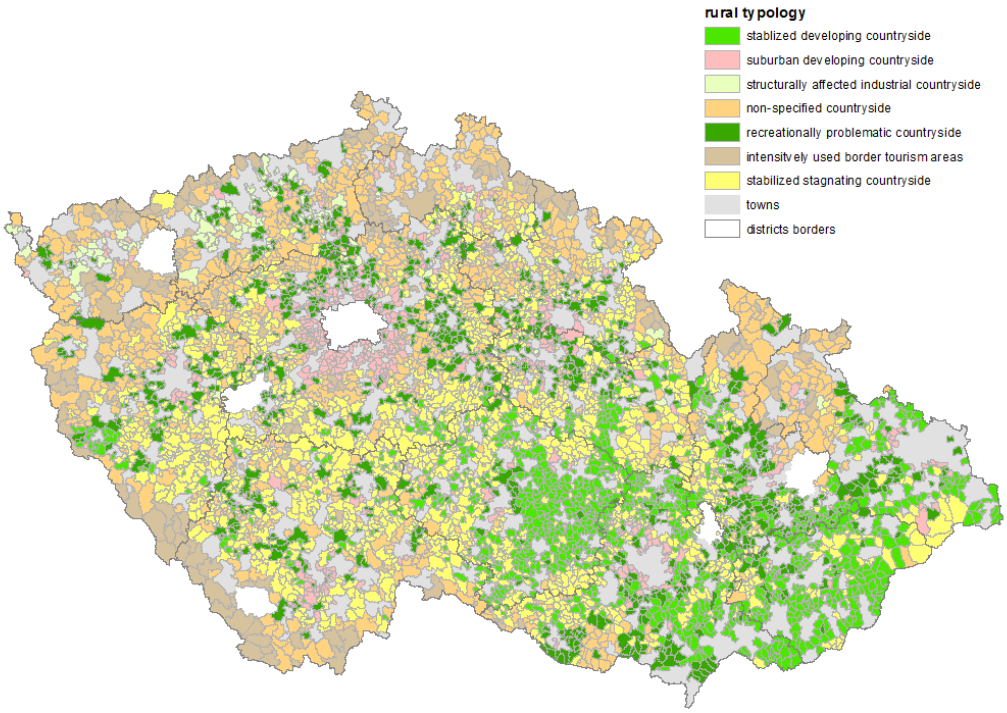


Figure 2 Typology of the Czech rural municipalities based on WSA

highest values of social stability and employment cluster are typical for this cluster. It corresponds with the Perlín's type called Equipped Moravian countryside to a certain amount. Cluster 2 (suburban countryside) has a low potential for economic and leisure activities, high social stability and employment and the highest value of utility function describing development. This cluster partially corresponds with Perlín's type "core countryside".

cluster (type of countryside)	size and importance	development	potential for economic and leisure activities	social stability and employment
1	0.0696	0.3758	0.2215	0.9990
2	0.0938	0.5064	0.1750	0.9603
3	0.1695	0.3631	0.1381	0.8219
4	0.0713	0.3387	0.3307	0.7779
5	0.0449	0.3651	0.2678	0.9980
6	0.0452	0.3426	0.2414	0.9115

Table 1 Average values of utility functions for clusters

Cluster 3 "structurally affected industrial countryside" reached the highest values of the utility function "municipality size and importance". It is typical by the lowest values of utility functions "potential for economic and leisure activities" and "social stability and employment" which are expected results. This type corresponds with Perlín's type of the same name. It was not identified in Klufová typology (see picture 1). Cluster 4 "intensively used border tourism areas" differs from other types in its highest value of the function "potential for economic and leisure activities" and lowest value of "social stability and employment". This cluster corresponds to Perlín's types called "recreationally problematic countryside" and "intensive recreational areas". Cluster 5 "stabilized stagnating countryside" consists of municipalities with the lowest values of their "size and importance" and low values of "development" on one hand, but with high values of "social stability and employment" on the other hand. The lowest values of "size and importance" and "development" functions are typical for "non-specified countryside".

4 Conclusion

The results of the statistical analyses conducted show that the basic differentiation of the Czech countryside is determined by the geographical location of particular areas (with the vertical position determined by the function and population size of a given settlement and the horizontal position determined by the proximity and accessibility of centres, as well as their macro-positional attractiveness, in terms of a west-east gradient). The spatial distribution of selected variables indicates that a variety of relationships exist among them. This confirms the affirmation that rural space in Czechia is not homogenous and that several types of rural space may be delimited, in terms of potential for development.

The presented differentiation of rural municipalities is based on the use of quantitative data to document current social-geographical phenomena in Czechias rural space. Although they only present the current situation in rural municipalities, the indicators used are influenced by long-term development trends. The typology presented should be further verified through qualitative research of internal development conditions of the countryside at the local level.

The typology computed on the base of WSA approach gave similar results as typologies based on various connections of multivariate statistical methods. It confirms existing identified types of the Czech rural space. Spatial distribution of particular rural types in typologies discussed above differs according to chosen variables and methods.

References

[1] Binek, et. al.: *Venkovský prostor a jeho oživení*. Georgetown, Brno. 2007

[2] Cloke, P., J.: An index of rurality for England and Wales. *Regional Studies* **11** (1977), 31–46

[3] Faltová Leitmanová, I. et. al.: Regioná lní rozvoj - přístupy a nástroje. Alfa Nakladatelství, Praha, 2012

[4] Green, R., J.: *Countryside Planning: The Future of Rural Regions*. Manchester University Press, Manchester, 1971

[5] Hampl, M.: *Realita, společnost a geografická organizace: hledání integrĺního řádu*. DemoArt, Praha, 1998

[6] Healey, M., J. and Ilbery, B., W.: *The Industralisation of the Countryside*. GeoBooks, Norwich, 1985

[7] Huigen, P. and Volkers, C., R.: Changes in rural areas in the Netherlands. In: *Limits to Rural Landuse* (Van Oort, G. M. R. A. et. al. eds.), , Pudoc. Wageningen, Amsterdam, 1989, 32–34

[8] Jablonský, J.: Software Support for Multiple Criteria Decision Problems. *Management Information Systems*, **4**, No. 2, (2009), 29–34

[9] Klufová, R., Rost, M. and Klicnarová, J.: *Modelování regionálních procesů*. Alfa Nakladatelství, Praha, 2012

[10] Klufová, R.: Vývoj a typologie českého venkova v kontextu prostorových souvislostí. Habilitation work (in preparation), University of South Bohemia, Faculty of Economics, České Budějovice, 2014

[11] Marler, R., T. and Arora, J., S.: The weigted sum method for multi-objective optimization: new insights. *Struct Multidisc Optim*, **41**, (2010), 853-862

[12] Perlín, R.: Countryside, a typology of rural areas. In.: *Česká etnoekologie* (Malý, F. and Viktoriová, B., eds.), Cargo Publisher, o. s., Praha, 1999

[13] Perlín, R., Kučerová, S. and Kučera, Z.: A Typology of Rural Space in Czechia according to its Potential for Development. *Geografie*, **115**, No. 2, (2010), 161-187

[14] Tzeng, G. and Huang, J.: *Multiple Attribute Decision Making: methods and applications*. CRC Press, London, 2011

Application of Adaptive Evolution Algorithm on real-world Flexible Job Shop Scheduling Problems

František Koblasa [1], František Manlig [2]

Abstract. Nowadays, scheduling of the manufacturing systems using heuristic algorithms; is a greatly debated, yet hardly ever used. There are several reasons such as complexity of available algorithms in the terms of setting up algorithm operators, the timespan of the optimization and suitability of the scheduling to the given manufacturing system.

This article concerns with the application of optimization methods on Flexible Job Shop Scheduling Problems (FJSSP) in the engineering small and medium enterprises (SME). There is briefly debated the way how SME are scheduling their manufacturing processes. Further, there is briefly described FJSSP model together with additionally used real-world constrains as setup with and without a necessity to have a part, transport and shift system. Developed Evolution Algorithm with adaptive procedure to set population size and number of generation is shown in the following section. Article also briefly describes developed methodology of assessing the appropriateness of the use of advanced heuristic methods in practice. Main focus of this article is given to the application of set methodology together with developed algorithm on two real world-examples.

Keywords: FJSSP, Evolution Algorithm, real-world scheduling problems, SME.

JEL Classification: C61
AMS Classification: 90C27

1 Introduction

Most of small and medium enterprises focus their effort mainly on downsizing their processes implementing methods of Lean Manufacturing. Nowadays, it is not enough and it is necessary to search for improvements in the other areas of the enterprise. One of these areas is the production planning and scheduling. SME are managing this area mainly by following ways:

- Manual planning and scheduling – mainly in the case of simple manufacturing systems.
- Planning boards and spreadsheets – frequently used especialy at the level of workshop as Scherer shows in [13] – It can be sufficient for small and simple manufacturing systems.
- The push systems as Enterprise Resource Planning - are based on the concept of Material Requirements Planning or Manufacturing Resource Planning. Despite the substantial benefits of these systems there is major problem of planning to unconstrained capacities.
- APS (Advanced Planning and Scheduling) – solves number of problems thanks to scheduling to the coinstrained capacities, thus only dispathing rules are used in the most cases.
- The pull systems based on Kanban – They focuses rather on management of the manufacturing system than to planning and scheduling as Thürer rewieved in [15].
- The hibrid systems as Drum – Buffer - Rope - are based on Theory Of Constraints – System is quiet efficient in the most cases, but it suffers phenomen of shifting bottleneck in the area of the job shop scheduling. Its principles were implemented in world known OPT system (Optimized Production Technology).
- Computer simulation – the main advantege is possiblity ability to simulate wide types of the manufacturing systems as Jobshops (flexible, hybrid) which can be seen in papers of Ottjes and Plechač [9,11]. Major disadvatege is than inability to generate Active Schedules as they use mostly only Non-delay Schedules.

The last possibility is than use of specialised optimization tool, which is using advanced heuristic algorithms such as Evolution Algorithms (EA). Those tools are hard to implement in the real world conditions. Main

[1] Department of Manufacturing Systems, Faculty of Mechanical Engineering, Technical University of Liberec, Studentská 2, Liberec 1, Czech Republic e-mail: frantisek.koblasa@tul.cz

[2] Department of Manufacturing Systems, Faculty of Mechanical Engineering, Technical University of Liberec, Studentská 2, Liberec 1, Czech Republic e-mail: frantisek.manlig@tul.cz

constain is complexity of the algorithm operator setup (e.g. size of population, number of generations, crossover propability etc.).

These issues will be discussed on the model Flexible Job Shop Scheduling Problems (FJSSP) in the following chapters. FJSSP model is described shortly at first. Developed Adaptive Evolution Algorithm implemented in the principle of cascade Automatic Optimization (AO) is than shown in the third chapter. Fourth chapter than showes used methodology for assessing the appropriateness of the implementation of AO in real-world conditions and describes the experience of testing the appropriateness of the implementation of cascade automatic optimization of the real-world models of two manufacturing companies.

2 Flexible Job Shop Scheduling Problem

Classical Job shop scheduling problem (JSSP) is the extension of Flow Shop Scheduling Problem described in [2] by Brucker and it is considered as the most common scheduling model [10] by Pinedo. The extension is in arbitrary trajectory of jobs through manufacturing system, besides Flow Shop Scheduling Problem, where all jobs has the same trajectory.

The most frequent modification of JSSP, especially in the engineering production where technology-oriented production design is used, is the Flexible Job Shop Scheduling Problem (FJSSP). FJSSP has to select the sequence of operations on each workshops and besides classical JSSP, it is necessary to assign each operation to particular machine in the workshop. FJSSP problem is JSSP extended by the assigning problem and could be described as follows. FJSSP problem has m machines and n jobs. Each job consists of a sequence of operation where they are allowed to be processed on any among a set of available machines, machine can only execute one operation at a given time. Preemption is not allowed.

JSSP and its flexible modification have wide range of application and its constraints beginning with the service as Tanev et al. shows in [14] ending with manufacturing as it is described in [12, 16]. There are lots of possible additional constraints as Demir [3] defined.

Theoretical FJSSP model was for our purpose extended by following constrains which are reflecting tested real world problems:

- Job release dates.
- Setup operations – with and without necessity to have part of preceding job to do the setup.
- Transport operations – Transport and manipulation between each workshops and machines.
- Shifts – multi shift systems (up to 3 period system – from one 8 hour shift up to two 12 hour shifts)

3 Adaptive Evolution Algorithm and Automatic Optimization

Designed EA follows classical structure of Simple Genetic Algorithm (initiation-selection-crossover – mutation) using Random Key (RK) representation developed by Bean [1]. Selection is made by roulette wheel principle, where the best solution has the highest probability to become parent and number of selected parents equals population size. Uniform crossover is used in reproduction stage with set probability. Instead of classical mutation principle, Critical Path Analysis based on Local Search is used, following schema presented by Nowicky et. al [8]. This approach is applied on defined percentage of selected parents, where critical block is randomly selected, the same as the first or the last pair of operation to be swapped.

Sometimes, cross over procedure cause, that RK used for sorting operation is damaged by genes, which have same RK, so algorithm is not able to determine their operation sequence. That is caused by "incest" cross over, where previously selected parent and its child pair to produce new individual. Incest control mechanism then repairs defect genes by calculating RK average of the last healthy RK gene and defect RK gene and replace defect random key gene by this average.

EA continues by selecting individuals (parents and children), by developed Elite strategy with Clone Control, to be allowed in to new generation. All individuals in elitist strategy (old generation and children) are selected and sorted by fitness function, than better half of individuals is allowed to proceed in to the new generation. This approach usually leads to fast convergence of the population to the best individual, so clone control mechanism is also applied. Clone control mechanism, after population sort, checks value of the individuals fitness function and if there is some individual with the same fitness (clone), than individual is replaced by individual that did not succeed in the selection. This procedure is much faster than comparing chromosome with chromosomes. However, it could lead to removing solutions that have the same fitness, but different sequence of operations (chromosome).

Fixed operators are than set as in [6]:

- Crossover probabity = 90%
- CPM Local search percentage = 1%
- Key operators as population size and number of generations are set by the mechanisms inside the automatic cascade optimization and are based on the remaining optimization timespan.

Principle of the Automatic Optimization (AO) is based on planner or workshop foreman need to get the best possible feasible schedule in the required time. Generally, there is a greater pressure on optimization timespan, than on fitness function. Workshop foreman needs some feasible schedule in given time more than the best schedule in the future.

AO begins with constructive algorithm(CA) using common dispatching rule as Shorter Processing Time (SPT), Longer Processing Time (LPT), Most Work Remaining (MWKR) and First-In First-Out (FIFO), knowing that this method give us feasible schedule fast. Based on SPT timespan, algorithm predicts the timespan of the following dispatching rule. If the predicted timespan is smaller than remaining optimization timespan, than the next schedule is calculated. Further predictions are based on average of the previous timespans. That principle compensates the different PC resource utilization over the time. Optimization continues with selecting schedules generated by dispatching rules and uses them as the initial solutions for single swap local search, as we present in [5], to get better results in the greater time than in the case of dispatching rules. Timespan of the optimization is checked each iteration of this method to predict further iteration. Algorithm also checks if required timespan did not exceed current time limit. EA is the last applied method which should give us the best results but in the longer timespan than previous methods.

Timespan of optimization is influenced by number of generation and by number of schedule calculation by CA. Number of generations can be easily determined by before defined timespan and timespan of CA runs with respect to number of individuals in generation.

Number of individuals in the generation can be set fixed or it is possible to use one of the adaptive population sizing schemes reviewed by Montero et al. [7]. We decided to use adaptive population sizing scheme, knowing that population size can have significant impact on explorative and exploitive function of EA.

We use of Population Resizing on Fitness Improvement GA principles (PRoFIGA) developed by Eiben et al. [4] due to its interesting results in the field of parameter less GA research. PRoFIGA resize the population by growing or shrinking based on an improvement of the best fitness contained in the population. Population grows when there is an improvement of the best fitness function or when there is no improvement in defined number of iteration (in our case 5) else population shrinks (decrease factor = increase factor). Growing factor X is than (1):

- *IF* –increase factor (0,1) – in our case 0.5 to get greedy behavior of resizing policy.
- *MN* –maximal number of generation, which in our case is initially set to 1000 and actualized every generation based on forecast of previous population number, its timespan calculation and overall remaining timespan.
- *CN* –number of current generation
- max*FITn* –current best fitness
- max*FITp* –previous best fitness
- *ini*max*FIT* –initial best fitness

$$X = IF * (MN - CN) * \frac{\max\ FITn - \max\ FITp}{ini\ \max\ FIT} \tag{1}$$

4 Methodology of assessing the appropriateness of the use of Automatic Optimization

It is necessary to take in account, same as during implementation APS, whole manufacturing system - not only machines, but also labors and present planning processes. That is the purpose of defined methodology, which its steps follows common rules given by Industrial Engineering or project management. Further chapters are following steps in defined methodology using real-world problems. There were used two examples of engineering companies whose models can be accurately described by FJSSP to test defined methodology and principles of AO.

Case A enterprise is focussing mainly on table type horizontal milling and boring machines. Request to make schedule optimization is based on the complicated planning of part of the manufacturing system, which is responsible to supply assembly workshop with milling heads. Main problem in this area are excessive working

queues, which are defined by term planning as a part of OPT system, which does not recognize this workshop as the bottleneck.

The second enterprise (Case B) is focusing on development, producing and installation of terminal units for medical gases distribution systems including fittings. Global goal of this case is to improve knowledge of conflict jobs (those which are requiring same machine in the same time) whose are not reflected in the present method of production planning (Material Requirements Planning).

4.1 Defining plane (schedule) requirements

Here, it is necessary to define all basic attributes, whose are influencing schedule creation and its final outcomes. Key attributes are plane evaluation criteria i.e. definition of global goals and recourses respectively constraints. The global goal of the both cases (A and B) is to minimize working queues at their respective machines.

4.2 Determination of measurable indicators - measuring global goal by the objective function

Determination of measurable indicators is based on the previous step of global goal definition. Those goals are quantified with respect to indicator monitoring in the practice.

Measurable indicators are can be divided as:

- Objective function – in the most cases objective function is defined by time factor as it is easily measurable in the practice.
- Manufacturing model indicators – Duration of technology operations (setting up the smallest time units for operations such as processing time, setup time etc.) a material and semi-finished pats unites (setting up cutting planes to convert material unites as tones to number of pieces etc.)

There were set seconds as simulation unites in both cases (A and B) and objective function of makespan. Seconds were set to calculate duration of the large manufacturing batches which consist of hundreds of small parts with less than minute processing times.

4.3 Current state analysis

Analysis of current state is the most time demanding step also with regard to usual necessity to change setup in the previous steps (define and measure) if they were set inappropriately. Analysis of current state can be divided in to following points:

- Select the main representatives: The main production of Case A is machining heads and light equipment. There were selected together with influenced laborers three types of machining heads which are produced most frequently with regard to different complexity of manufacturing process. There were selected all job demands, in the Case B with regard to wide product base, which release dates are within same week. Model is completed with make to stock products selected thanks to ABC analysis.
- Manufacturing system type: Both of manufacturing systems (Case A and B) have strong indices of FJJSP such as different job trajectories through manufacturing system, jobs are frequently returning to stages where they were already processed and workshops designs are technology oriented.
- Number of jobs i.e. tasks and its duration: The average processing time takes about 1900 minutes, setup time about 700 minutes and manufacturing batches are about 1-10 pieces in the Case A. The Case B has wide range of both processing and setup time duration (from second up to 30 minutes) and it has wide variety of manufacturing batches (1-1000 pcs).
- Data consistency: Transport times (in both cases) are given by distance and expert estimation by manipulation labourers. Processing and setup times are given by technology orders inside information system . There were several revisions of these data before they were used to make model. An example could be control operation which has both processing time and setup time equal to zero due cost policy. Most inaccuracies in the Case B were than operation bounded with cooperation, where deviation before and after revision were 3-7 days.
- Constraints: Both cases have general constraints as machines, type of setup, shift system. Additional constraints as transport, cooperation or technologies like annealing furnaces, blackening etc. are scheduled as non-constrained capacities by term planning. Case A model consists of 35 machines which are processing 80 job demands (each has 10 operation in average). Case B has 101 machines which are processing 229 job demands (each has 5 operations in average).
- Used planning methods: Planning methods differs a lot from case to case. Company of Case A is using APS system based on OPT and company of Case B is using Material Requirements Planning . That also influence

readiness to use operation based scheduling as workers of case A are suit to schedule bottleneck and workers of case B are suite only to use master production schedule.

- Finished production evidence system: Both cases are using automatic identification systems (bar codes) so there are no significant delays between finishing operation on product and its evidence in the system.
- Planning and scheduling time demand: Rescheduling is made in the both cases over night shift, so there is plenty of time for optimization. Nevertheless, we test optimization timespan of 15 minutes with regard to maximal available time to reschedule due random events like constraint unavailability (e.g. machine).

4.4 Testing models by developer scheduling methods

Model testing is done in the three phases. Verification of data consistency, model ability to provide real schedules and analysis of timespan-number of scheduled operations dependency is verified in the first phase. Timespan-operations dependency is only tentative because it does not usually take into account the combinatorial complexity of final model. Thanks to this test it was found, in the case A and B, that there are several problems in the data consistence, mainly in unrealistic setup of release and due dates – sum of manufacturing batches duration was longer than difference between release and due dates.

The following test (phase two) contains additional one third of final model. This model is used to analyse timespan and combinatorial complexity. This test can also preview possible benefits (in the terms of objective function), of implemented optimization methods thanks to comparing results of models with and without material release dates. Model without release dates can show further dependence of benefits (makespan) and optimization timespan. Model with release dates can show, if set due dates are real, comparing schedule results with those given by recent used planning method (Case A – OPT).Case models were tested by CA and EA method, LS was skipped because of forecasted timespan given by CA. Combinatorial complexity of both cases rise as schedules given by dispatching rule RS shown us different makespan values.

The last phase contains whole planning horizon and suit to verify ability of advanced optimization methods to achieve best values of objective functions using real manufacturing model with all defined constraints. This model is also used to validate potential benefits with regard to frequency of checking-out finished operations i.e. products (using objective function of makespan or job tardiness).

Difference of objective function values generated by RS (about 15% see Table 1 RS_b – best makespan, RS_w worst and RS_a .average of 10 measurements) shows significant combinatorial complexity of Case A. Difference between RS generated values in the Case B (so complexity) is a bit lower (about 5%). Comparing Case A used scheduling policy (Non delay schedules using FIFO) with result given by EA we can see possible improvement of makespan (about 10% = 6 days). Using same comparison for Case B we can reduce makespan by approximately 14% which means 5 days makespan reduction.

		Case A		Case B	
Method	Dispatching rule	Active [min]	Non Delay [min]	Active [min]	Non Delay [min]
CA	SPT	109206	78666	57130	55684
	LPT	81846	91105	59020	62414
	MWKR	80406	80260	55724	63574
	FIFO	78876	83865	59019	59998
	RS_b	78966	81195	52139	64804
	RS_w	93366	91836	55685	70565
	RS_a	83795	85211	54185	67129
EA		74892		51930	

Table 1 Final models makespan results

5 Conclusion

The most important step implementing methods as AO is current state analysis. This step can, apart from analysis of common constraints (machines, release dates etc.), show fundamental constraints which are hard to schedule by classical Job Shop Scheduling. The example could be Case A model which contains chemical-physical operations as annealing or blackening. Manufacturing batches of those operations are merging and dividing base on the specific condition of each technology (e.g. works space of annealing furnace) so it is not

possible to use classical batch scheduling. Case B showed another type constraint – external cooperation with wide range of due dates (3-7 days divergence). Those constraints (batches) were scheduled by term planning using operation of transport with maximal evidenced duration.

AO can be implemented in both cases, as they are suitable for FJSSP and they have certain level of combinatorial complexity, but with several limitations. It is necessary to make revision of technological order data, release and due date policy. Moreover, it is necessary to standardize supply delivery dates and to implement Manufacturing Resource Planning methods before implementation of AO, in the case B, to make company suitable to define processing and setup times properly. Nevertheless, tests of AO show that potential implementation benefit could be significant (10% Case A, 15% Case B reduction of makespan) comparing recently used planning policy and developed EA.

Further research will be focused on improving scheduling of operations as annealing combining classical sequencing together with combinatorial problems as Knapsack or Bin-Packing problem.

Acknowledgements

Supported by the grant No. SGS-21010 of the Grant Agency TUL-SGS.

References

[1] Bean, J. C.: Genetic Algorithms and Random Keys for Sequencing and Optimization, *ORSA Journal on Computing*, vol.6, no.2, Spring 1994

[2] Brucker, P.: *Scheduling Algorithms*. 5th edition. [s.l.]: Springer, 2007. 383 s. ISBN 978-3540695158, 354069515X.

[3] Demir, Y. and İşleyen, S.K.: Evaluation of mathematical models for flexible job-shop scheduling problems, *Applied Mathematical Modelling*, Volume 37, Issue 3, 1 February 2013, Pages 977-988, ISSN 0307-904X, http://dx.doi.org/10.1016/j.apm.2012.03.020.

[4] Eiben, A. E., Marchiori, E. and Valkó V. A.: Evolutionary algorithms with on-the-fly population size adjustment. In *Parallel Problem Solving from Nature*, PPSN VIII, volume 3242 of Lecture Notes in Computer Science. Springer, 2004, 41–50

[5] Koblasa, F., Dias, L. S., Oliveira, J. A. and Pereira, G.: Heuristic Approach as a way to Improve Scheduling in ERP/APS Systems. In: *Proceedings of 15th European Concurrent Engineering Conference (ECEC2008)*. Eds. A. Brito and J.M. Teixeira, Porto. EUROSIS-ETI Publication. ISBN 978-9077381-399-7, 2008. 47-51,

[6] Koblasa, F., Manlig, F. and Vavruška, J.: Evolution algorithm for job shop scheduling problem constrained by the optimization timespan, *Applied Mechanics and Materials*, vol. 309, 2013, 350-357.

[7] Montero, E. and Riff, M. C.: On-the-fly calibrating strategies for evolutionary algorithms, *Information Sciences*, Volume 181, Issue 3/1, ISSN 0020-0255, 10.1016/j.ins.2010.09.016, (2011), 552-566.

[8] Nowicky, E. and Smutnicki, C.: A fast tabu search algorithm for job shop problem. *Manag. Sci*., vol. 42, 1996 797–813.

[9] Ottjes, J.A. and Veeke, H. P. M.: Production Scheduling of Complex Jobs with Simulation. In: *Proceedings of the Business and Industry Simulation Symposium (ASTC 1999)*. Washington D.C. [SCS]. April 2000. ISBN 1-56555-199-0

[10] Pinedo, M. L: *Scheduling: Theory, Algorithms, and Systems*, Third Edition, Springer, 2008

[11] Plechač, R., Manlig, F. and Koblasa, F.: Job shop scheduling with aid of computer simulation. *University Review* Vol. 4, No. 2. Trenčín: Alexander Dubček University of Trenčín, 2010, s. 29-33 ISSN 1337-6047

[12] Rui, Z. and Cheng, W.: A hybrid local search algorithm for scheduling real-world job shops with batch-wise pending due dates, *Engineering Applications of Artificial Intelligence*, Volume 25, Issue 2, March 2012, Pages 209-221, ISSN 0952-1976, 10.1016/j.engappai.2011.04.014.

[13] Scherer, E.: Results of the ERP user satisfaction survey: Conclusion for the real-world scheduling, In: *7th International Workshop on Human Factors in Planning, Scheduling and Control in Manufacturing*, 2005

[14] Tanev, I. T, Uozumi, T. and Morotome, T.: Hybrid evolutionary algorithm-based real-world flexible job shop scheduling problem: application service provider approach, *Applied Soft Computing*, Volume 5, Issue 1, 2004, Pages 87-100, ISSN 1568-4946

[15] Thürer, M., Land, M. J. and Stevenson, M.: Card-based workload control for job shops: Improving CONWIP-BACABANA, *International Journal of Production Economics*, Volume 147, Part A, 2014, 180-188, ISSN 0925-5273, http://dx.doi.org/10.1016/j.ijpe.2013.09.015.

[16] Yusof, R., Khalid, M. and San, T.CH.: Solving industrial based job-shop scheduling problem by distributed micro-genetic algorithm with local search. In: Proceedings of the 14th international conference on Knowledge-based and intelligent information and engineering systems: Part I, (KES'10), Springer-Verlag, Berlin, Heidelberg, 2010 69-79.

Fuzzy approach for portfolio selection problem

Maria Kobzareva[1], Jan Pelikán[2]

Abstract. Fuzzy approach is a recent approach to uncertainty modeling, which deals with uncertainty by describing it using fuzzy sets defined by fuzzy membership functions. Fuzzy approach is a modeling tool, which solves problems in which data are uncertain. This problem is frequently solved problem in many fields and it is more appropriate for real-life applications, because in practice data are hardly deterministic, but tend to change unexpectedly. This paper introduces fuzzy decomposition methodology and ranking function approach to solve fuzzy defined problems and suggests a possibility of solving portfolio optimization problem with uncertain data set using presented methods. Fuzzy portfolio optimization problem takes into account the fact that share profits could not be the same as predicted, but real share profits values could be different; the future profit of the shares are presented in the model by triangle fuzzy numbers. In this paper classic portfolio optimization problem is transformed into fuzzy linear programming problem and solved by presented techniques. Results obtained by fuzzy approaches used in this paper are compared with *p*-fractile approach which is briefly reviewed in the paper. Presented applications are described on a case study and provided with mathematical models and its detailed description. The paper also presents computation experiments and its results.

Key words: bound and decomposition method, data uncertainty, fuzzy linear programming, fuzzy variables, *p*-fractile approach, portfolio optimization problem, ranking function, triangle fuzzy number

JEL Classification: C6, G1

1 Introduction

One of the recent approaches to uncertainty modeling is fuzzy set approach, where set is described by fuzzy membership function. We distinguish two main classes of fuzzy problems: problems with fuzzy variables and problems with fuzzy parameters. The problems in which both above mentioned elements are applied are called fully fuzzified linear programming (FFLP) problems. Fully fuzzy problem is therefore generalized version of fuzzy linear programming problem where all decision parameters and variables are fuzzy numbers.

This paper presents portfolio problem formulated as fuzzy problem with fuzzy parameters and suggests two approaches to solve the problem. Bound and decomposition method and ranking function method are introduced to find an optimal solution for fuzzy linear programming (FLP) portfolio optimization problem. In the bound and decomposition method the given FLP problem is decomposed into three crisp linear programming (CLP) problems with bounded variables constraints, the three CLP problems are solved separately and by using its optimal solutions, the fuzzy optimal solution to the given FLP problem is obtained. Fuzzy ranking function method calculates adjusted profit values using ranking function for triangle fuzzy numbers and by solving basic portfolio optimization model with ranking profits optimal solution is obtained. The paper also reviews *p*-fractile approach used to solve portfolio optimization problem with uncertain data. The methods are illustrated on a case study, results are provided.

The paper is organized as follows: in section 2 basic formulation of portfolio optimization model is introduced, section 3 contains preliminaries for fuzzy numbers and operations with fuzzy numbers. Section 4 describes fuzzy linear programming portfolio problem. Section 5 is associated with bound and decomposition method, in section 6 ranking function method and the properties of the presented linear ranking function are introduced. Section 7 relates to the *p*-fractile approach to portfolio selection problem. In section 8 fuzzy portfolio optimization problem is solved using the above mentioned methods, results are presented and compared. Section 9 briefly summarizes presented techniques and results.

[1]University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill sq.4, Prague 3, Czech Republic, e-mail: mariakobzareva@gmail.com

[2] University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill sq.4, Prague 3, Czech Republic, e-mail: pelikan@vse.cz

2 Portfolio selection problem and its mathematical formulation

Let me introduce portfolio optimization problem, where we have to select shares from a given number of shares, respecting the condition that the value of selected shares will not exceed given investment limit *K*. The aim is to maximize future value of selected shares, future values of share profits are given. The problem is that the real profit value may be different from the predicted value. The purpose of the case study is to find so called fuzzy optimal solution under given circumstances, which respects unexpected decrease or increase in future share profits values.

Basic portfolio optimization problem model can be formulated as follows: let us define *n* as a given number of shares from which we should select a subset, *K* is the limit of total expenses for selected shares, for share *i* a_i is the price of share, c_i is expected profit of share *i*. Binary variable x_i defines the decision to choose share *i*, when x_i=1, or decision not to choose the share in case x_i=0[3].

The basic model of choosing the shares is described below:

$$max \sum_{i=1}^{n} c_i x_i \,, s.t. \sum_{i=1}^{n} a_i x_i \leq K, x_i \in \{0,1\}\; i = 1,2, \dots, n. \tag{1}$$

According to this model we will create a fuzzy model in the following section.

3 Preliminaries

We will use triangle fuzzy number $\tilde{a} = (\underline{a}, \hat{a}, \bar{a})$ (see Figure 1), where $\underline{a} \leq \hat{a} \leq \bar{a}$ and membership function $\mu_{\tilde{a}}(x)$ is described as follows:

$$\mu_{\tilde{a}}(x) = \begin{cases} 0, & x < \underline{a}, \\ L(x) = \frac{x-\underline{a}}{\hat{a}-\underline{a}}, & \underline{a} \leq x \leq \hat{a}, \\ R(x) = \frac{\bar{a}-x}{\bar{a}-\hat{a}}, & \hat{a} \leq x \leq \bar{a}, \\ 0, & x > \bar{a}. \end{cases} \tag{2}$$

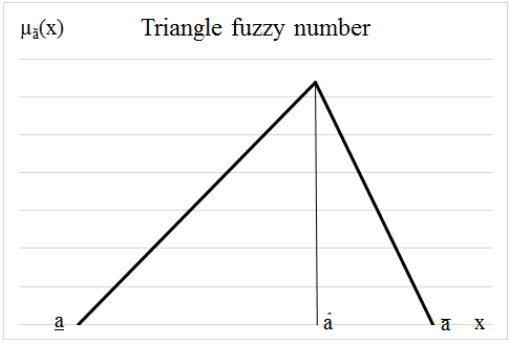


Figure 1: Triangle fuzzy number

Let's now define summation (3) and multiplication (4) for triangle fuzzy numbers: let $\tilde{a} = (\underline{a}_i, \hat{a}_i, \bar{a}_i)$ and $\tilde{b} = (\underline{b}_i, \hat{b}_i, \bar{b}_i)$ be triangle fuzzy numbers, then

$$\tilde{a} + \tilde{b} = (\underline{a} + \underline{b}, \hat{a} + \hat{b}, \bar{a} + \bar{b}), \tag{3}$$

$$k * \tilde{a} = (k * \underline{a}, k * \hat{a}, k * \bar{a})\; pro\; k > 0. \tag{4}$$

4 Fuzzy linear programming portfolio optimization problem

First let's define fuzzy relations the proposed below method is based on. Let $\tilde{a} = (\underline{a}_i, \hat{a}_i, \bar{a}_i)$ and $\tilde{b} = (\underline{b}_i, \hat{b}_i, \bar{b}_i)$ be triangle fuzzy numbers, then

$$\tilde{a} \approx \tilde{b} \text{ if and only if} (\underline{a} = \underline{b}, \hat{a} = \hat{b}\; and\; \bar{a} = \bar{b}), \tag{5}$$

$$\tilde{a} \preccurlyeq \tilde{b} \text{ if and only if} (\underline{a} \leq \underline{b}, \hat{a} \leq \hat{b}\; and\; \bar{a} \leq \bar{b}), \tag{6}$$

$$\tilde{a} \succcurlyeq \tilde{b} \text{ if and only if } (\underline{a} \geq \underline{b}, \hat{a} \geq \hat{b}\; and\; \bar{a} \geq \bar{b}). \tag{7}$$

[3] Presented model is designed for estate market, where you can either purchase estate or not. In the paper in order to simplify the model we work with share market.

Consider the following fuzzy linear programming portfolio optimization problem with n fuzzy variables, formulated as follows:

$$max \sum_{i=1}^{n} \tilde{c}_i \tilde{x}_i \, , s.t. \sum_{i=1}^{n} \tilde{a}_i \tilde{x}_i \leq \tilde{K}, \tilde{x}_i \in \{0,1\} \; i = 1,2, \dots , n. \tag{8}$$

Let the parameters $\tilde{a}_i$, $\tilde{c}_i$ and $\tilde{K}$ be the triangular fuzzy number ($\underline{a}_i$, $\hat{a}_i$, $\bar{a}_i$), ($\underline{c}_i$, $\hat{c}_i$, $\bar{c}_i$) and ($\underline{K}$, $\hat{K}$, $\bar{K}$) respectively. Then, the problem (1) can be formulated as follows [5, 249]:

$$max(z_1, z_2, z_3) \approx \sum_{i=1}^{n} (\underline{c}_i, \hat{c}_i \, , \bar{c}_i) * (\underline{x}_i, \hat{x}_i, \bar{x}_i), s.t. \sum_{i=1}^{n} (\underline{a}_i, \hat{a}_i, \bar{a}_i) * (\underline{x}_i, \hat{x}_i, \bar{x}_i) \preccurlyeq (\underline{K}, \hat{K}, \bar{K}) \, , \tag{9}$$

where $(\underline{x}_i, \hat{x}_i, \bar{x}_i) \in \{0,1\}$ for $i = 1,2, \dots , n$ are fuzzy numbers and the following constraint is fulfilled:

$$0 \leq \underline{x}_i \leq \hat{x}_i \leq \bar{x}_i, \; i = 1, 2, \dots, n. \tag{10}$$

Using the arithmetic operations and partial ordering relations let's decompose the given FLPP as follows [5, 249]:

$$max \; z_1 = lower \; value \; of \sum_{i=1}^{n} (\underline{c}_i, \hat{c}_i, \bar{c}_i) * (\underline{x}_i, \hat{x}_i, \bar{x}_i)$$

$$max \; z_2 = middle \; value \; of \sum_{i=1}^{n} (\underline{c}_i, \hat{c}_i, \bar{c}_i) * (\underline{x}_i, \hat{x}_i, \bar{x}_i)$$

$$max \; z_3 = upper \; value \; of \sum_{i=1}^{n} (\underline{c}_i, \hat{c}_i, \bar{c}_i) * (\underline{x}_i, \hat{x}_i, \bar{x}_i)$$

Subject to (11)

$$\sum_{i=1}^{n} lower \; value \; of \; (\underline{a}_i, \hat{a}_i, \bar{a}_i) * (\underline{x}_i, \hat{x}_i, \bar{x}_i) \leq \underline{K}_i \, , for \; i = 1,2, \dots , n;$$

$$\sum_{i=1}^{n} middle \; value \; of \; (\underline{a}_i, \hat{a}_i, \bar{a}_i) * (\underline{x}_i, \hat{x}_i, \bar{x}_i) \leq \hat{K}, for \; i = 1,2, \dots , n;$$

$$\sum_{i=1}^{n} upper \; value \; of \; (\underline{a}_i, \hat{a}_i, \bar{a}_i) * (\underline{x}_i, \hat{x}_i, \bar{x}_i) \leq \bar{K}, for \; i = 1,2, \dots , n;$$

and all decision variables are non-negative.

From the above decomposition problem, we construct the following CLP problems, namely middle level problem (MLP), upper level problem (ULP) and lower level problem (LLP) as follows:

(MLP)

$$max \; z_2 = \sum_{i=1}^{n} \hat{c}_i * \hat{x}_i$$

subject to (12)

$$\sum_{i=1}^{n} \hat{a}_i * \hat{x}_i \leq \hat{K}, for \; i = 1,2, \dots , n; \; \hat{x}_i \in \{0,1\} \; i = 1,2, \dots , n.$$

At least one decision variable of the (MLP) should occur in the model and all decision variables should be non-negative.

(ULP)

$$max \; z_3 = \sum_{i=1}^{n} \bar{c}_i * \bar{x}_i$$

subject to (13)

$$\sum_{i=1}^{n} \bar{c}_i * \bar{x}_i \geq z_2,$$

$$\sum_{i=1}^{n} \bar{a}_i * \bar{x}_i \leq \overline{K}_i, for \; i = 1,2, \dots , n; \; \bar{x}_i \in \{0,1\} \; i = 1,2, \dots , n.$$

In decomposition (ULP) problem at least one decision variable of the (ULP) should occur that is not used in (MLP); all variables in the constraints and objective function in (ULP) must satisfy the bounded constraints; replacing all values of the decision variables which are obtained in (MLP) and all decision variables are non-negative, where z_2^* is the optimal objective value of (MLP).

and (LLP)

$$max \; z_1 = \sum_{i=1}^{n} \underline{c}_i * \underline{x}_i$$

subject to (14)

$$\sum_{i=1}^{n} \underline{c}_i * \underline{x}_i \leq z_2,$$

$$\sum_{i=1}^{n} \underline{a}_i * \underline{x}_i \leq \underline{K}, for \; i = 1,2, \dots , n; \; \underline{x}_i \in \{0,1\} \; i = 1,2, \dots , n.$$

Again, at least one decision variable of the (LLP) should occur which is not used in (MLP) and (ULP); all variables in the constraints and objective function in (LLP) must satisfy the bounded constraints; replacing all values of the decision variables which are obtained in the (MLP) and (ULP) and all decision variables are non-negative, where z_2^* is the optimal objective value of (MLP).

Remark 4.1: In the particular case of portfolio optimization we only consider share profits to be fuzzy numbers, share prices and investment capital are fixed values at the moment we purchase shares. In the model above then for all shares i $\underline{a}_i = \hat{a}_i = \bar{a}_i$ and $\underline{K} = \hat{K} = \bar{K}$.

5 Bound and decomposition method

This section describes bound and decomposition method for solving a FLP problem presented in [5, 250]. The algorithms to solve fuzzy portfolio optimization model with bound and decomposition method proceeds as follows.

Step 1: Construct (MLP) (12), (ULP) (13) and (LLP) (14) problems from the given FLP portfolio optimization problem.

Step 2: Using existing linear programming technique, solve the (MLP) problem, then the (ULP) problem and then the (LLP) problem in the order only and obtain the values of all real decision variables $\underline{x}_i$, $\hat{x}_i$ and $\bar{x}_i$ and values of all objectives z_1, z_2 and z_3. Let the decision variables values be $\underline{x}_i^*$, $\hat{x}_i^*$ and $\bar{x}_i^*$ $i = 1, 2, \ldots, n$ and objective values be z_1^*, z_2^* and z_3^*.

Step 3: An optimal fuzzy solution to the given FLP problems is $\tilde{x}_i^* = (\underline{x}_i, \hat{x}_i, \bar{x}_i)$, $i=1, 2, \ldots, n$ and the maximum fuzzy objective is $\tilde{z}^* = (z_1^*, z_2^*, z_3^*)$. (by the Theorem 4.1.[5, 250]).

6 Ranking function approach

An efficient approach for ordering the elements is to define a ranking function D: $F(\Re) \rightarrow \Re$ which maps each fuzzy number into the real line, where a natural order exists. Let's define orders on by [1, 24]: let $\tilde{a} = (\underline{a}_i, \hat{a}_i, \bar{a}_i)$ and $\tilde{b} = (\underline{b}_i, \hat{b}_i, \bar{b}_i)$ be triangle fuzzy numbers, then

$$\tilde{a} \succcurlyeq \tilde{b} \text{ if and only if } D(\tilde{a}) \geq D(\tilde{b}), \tag{15}$$

$$\tilde{a} \preccurlyeq \tilde{b} \text{ if and only if } D(\tilde{a}) \leq D(\tilde{b}), \tag{16}$$

$$\tilde{a} \approx \tilde{b} \text{ if and only if } D(\tilde{a}) = D(\tilde{b}). \tag{17}$$

Let's now restrict our attention to linear ranking function – for any triangular fuzzy number $\tilde{a}$ let's use ranking function D as follows [1, 24]:

$$D(\tilde{a}) = \frac{1}{2} * \left(\int_0^1 L(r)^{-1} dr + \int_0^1 R^{-1}(r) dr \right), \tag{18}$$

$$D(\tilde{a}) = \hat{a} + 1/4(\bar{a} + \underline{a} - 2 * \hat{a}). \tag{19}$$

Then, for triangular fuzzy number $\tilde{A}$ and $\tilde{B}$, we have:

$$\tilde{a} \succcurlyeq \tilde{b} \text{ if and only if } \hat{a} + 1/4(\bar{a} + \underline{a} - 2 * \hat{a}) \geq \hat{b} + 1/4(\bar{b} + \underline{b} - 2 * \hat{b}). \tag{20}$$

In this paper we use the presented ranking method to calculate future share profit values.

The method proposed in [1] suggests a fuzzy formulation of model which gives us fuzzy solution. The disadvantage of fuzzy solution is the fact, that we obtain the solution x as a triangle fuzzy number, which doesn't give us the idea, which shares to select, as there are always three values for each share. To avoid the problem which was mentioned above in ranking function method we can use the simplified version of the model, in which we assume, that the only parameter, presented by a triangle fuzzy number is the future profit of the shares, as it is the only parameter that is not known at the moment we choose to buy the shares. In this model we use ranking function to obtain the future profit value and then solve the classic portfolio problem (1), using ranking function coefficients for the share profit values. Another possibility is to use ranking function to obtain ranking prices in order to select the shares according to possible future changes in share prices along with ranking profit values in one model. In this paper we only assumed the share profits to be fuzzy numbers.

The problem can then be written as

$$max \sum_{i=1}^{n} c_i^r x_i \, , s.t. \sum_{i=1}^{n} a_i x_i \leq K, x_i \in \{0,1\} \; i = 1,2, \ldots \, , n, \tag{21}$$

where c_i^r is calculated according to (19).

7 Possibility *p*-fractile approach

Another approach, used to solve problems with uncertain parameters is *p*-fractile approach, described in [4, 11].

Let $\tilde{a}$ be a fuzzy number and g a real number. Now we define the measure of possibility of the event as $\alpha \in \tilde{a}$, which is described in the interval $(-\infty, g >$ by

$$Pos(\alpha \leq g) = sup_{r \in (-\infty, g)} \mu_{\tilde{a}}(r) \, . \tag{22}$$

The necessity of event that $\alpha \in \tilde{a}$ is certain in the interval $(-\infty, g >$ is then described by

$$Nes(\alpha \leq g) = 1 - sup_{r \in (-\infty, g)} \mu_{\tilde{a}}(r). \quad (23)$$

Similarly we can define $Pos(\alpha \geq g)$ and $Nes(\alpha \geq g)$. Now let's define *p*-necessity fractile. *P*-necessity fractile is given for $p \in (0,1)$: *p*-necessity fractile $u_p = \max u$, where *u* satisfies: $Nes(\alpha \geq u) \geq p$.

Now we define the portfolio problem based on *p*-fractile. For fuzzy $\tilde{c}_i$ and binary variable x_i the objective function $\tilde{z}(x) = \sum_{i=1}^{n} \tilde{c}_i x_i$ is a fuzzy number calculated according to (4) as

$$\tilde{z}(x) = \sum_{i=1}^{n} \tilde{c}_i x_i = (\sum_{i=1}^{n} \underline{c}_i * x_i, \sum_{i=1}^{n} \hat{c}_i * x_i, \sum_{i=1}^{n} \bar{c}_i * x_i). \quad (24)$$

Let's set $p \in (0,1)$ to *p*=0,8, than we can formulate portfolio optimization problem as a maximization of $u^{0,8}$, where $u^{0,8} = max(u)$ for $Nes(z \geq u) \geq 0{,}8$. From figure 2 it's clear that $u^{0,8} = \hat{z} - 0{,}8(\hat{z} - \underline{z})$.

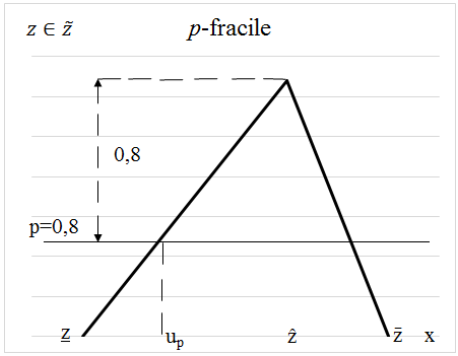


Figure 2: *p*-fractile

The problem can then be formulated as follows:

$$max \sum_{i=1}^{n} c_i^{0,8} x_i, s.t. \sum_{i=1}^{n} a_i x_i \leq K, x_i \in \{0,1\}\ i = 1,2, \dots, n, \quad (25)$$

where $c_i^{0,8}$ is calculated as $Nes(c_i \geq c_i^{0,8}) \geq 0{,}8, c_i \in \tilde{c}_i$.

8 Case study

Let us solve the problem with investment limit *K*=500 and a set of 30 shares, from which a subset has to be selected. The data subset was obtained from web pages http://finance.yahoo.com/, where we used path "investment, stocks, earning dates" to collect necessary data. Each share has a current price, which was stated in the beginning of April 2014, estimated profit of the share – low, average and high profit $\underline{c}_i$, $\hat{c}_i$ and $\bar{c}_i$ respectively. The price a_i of share *i*, the profit $\underline{c}_i$, $\hat{c}_i$ and $\bar{c}_i$ are given in the table 1.

	SIRI	FB	BAC	AA	PFE	CSCO	F	PBR	INTC	MU
$\underline{c}_i$	0,08	1,07	0,95	0,23	2,13	1,96	1,13	1,32	1,7	2,45
$\hat{c}_i$	0,1	1,26	1,08	0,43	2,23	1,99	1,34	1,75	1,86	3,02
$\bar{c}_i$	0,14	1,49	1,20	0,72	2,31	2,06	1,45	2,52	2,12	3,38
a_i	3,14	62,4	16,6	13	31,23	23,1	15,84	13,8	27	21,7
	GE	**T**	**GM**	**YHOO**	**GRPN**	**VZ**	**ZNGA**	**RAD**	**MSFT**	**GILD**
$\underline{c}_i$	1,65	2,52	3,1	1,13	-0,03	3,32	-0,04	0,33	2,6	2,85
$\hat{c}_i$	1,7	2,7	3,58	1,57	0,11	3,51	0,01	0,38	2,7	3,99
$\bar{c}_i$	1,78	2,85	4,05	2,03	0,22	3,62	0,04	0,46	2,81	5,79
a_i	25,58	35,1	33,3	33,4	7,06	47,5	4,07	6,94	39,4	65,5
	JPM	**EBAY**	**C**	**ORCL**	**HPQ**	**WFC**	**HBAN**	**RF**	**KO**	**CONN**
$\underline{c}_i$	5,12	2,91	4,1	2,88	3,54	3,75	0,68	0,78	2	3,38
$\hat{c}_i$	5,7	2,99	4,71	2,91	3,71	4,07	0,72	0,85	2,09	3,52
$\bar{c}_i$	6,12	3,06	5,8	2,93	3,83	4,35	0,76	0,92	2,13	3,73
a_i	57,4	54,1	46,2	39,8	32,8	47,7	9,42	10,3	40,2	45,4

Table 1: List of shares and their parameters

We had used the three approaches described in this paper and solve the presented linear problem, results are shown in the table 2. Beside the optimal value F in the table you can also find the corresponding solution *x*, which denotes

the share selection, the pessimistic estimate of profits for presented solution $\sum_{i=1}^{n} \underline{c}_i$, average estimated profit of solution $\sum_{i=1}^{n} \hat{c}_i$, and optimistic estimate profit for the solution $\sum_{i=1}^{n} \bar{c}_i$.

method	Solution x	F	$\sum \underline{c}_i$	$\sum \hat{c}_i$	$\sum \bar{c}_i$
MLP	1 0 1 0 0 1 1 1 0 1 0 1 1 0 0 1 0 0 0 0 1 0 1 1 1 1 1 1 0 1	45,26	41,06	45,26	49,71
LLP	1 0 1 0 0 1 1 1 0 1 0 1 1 0 0 1 0 0 0 0 1 0 1 1 1 1 1 1 0 1	41,06	41,06	45,26	49,71
ULP	1 0 1 0 0 1 1 1 0 1 0 1 1 0 0 1 0 0 0 0 1 0 1 1 1 1 1 1 0 1	49,71	41,06	45,26	49,71
Ranking	0 0 0 0 1 1 1 1 1 1 0 1 1 0 0 1 0 0 0 0 1 0 1 0 1 1 1 1 0 1	47,48	40,98	45,26	49,87
p-fractile	0 0 0 0 0 1 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 1 1 1 1 0 0 0 1	41,98	41,17	45,21	49,5

Table 2: Optimal solution

As you can see in table 2 maximum objective function value for decomposition method is F = 41,06 with the corresponding share selection x = (1 0 1 0 0 1 1 1 0 1 0 1 1 0 0 1 0 0 0 0 1 0 1 1 1 1 1 1 0 1), which is identical for all three models; for ranking function model the objective value is F = 47,48, corresponding share selection x = (0 0 0 0 1 1 1 1 1 1 0 1 1 0 0 1 0 0 0 0 1 0 1 0 1 1 1 1 0 1); for *p*-fractile approach optimal share set is described as follows x = (0 0 0 0 0 1 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 1 1 1 1 0 0 0 1) with the maximum value of objective function F = 41,98. In the table above you can also find and compare profits in case of pessimistic, modal or optimistic situation on the market.

As you can see from the table above, optimal solution for presented approaches is different. Using decomposition method we would select shares 1, 3, 6 to 8, 10, 12, 13, 16, 21, 23 to 28 and 30 total average expected profit would be 45,26, on the other hand if we use ranking function method we would obtain the following solution – shares 5 to 10, 12, 13, 16, 21, 23 25 to 28 and 30 would be selected with the same amount of total average expected profit – 45,26. If we compare those results with the solution, obtained from solving the same problem with *p*-fractile approach, where we select shares 6 to 8, 10, 12, 13, 16, 19, 21, 23 to 26 and 30, we can observe that the value of objective function achieved by *p*-fractile approach is slightly less in optimistic and modal situation but a little higher in case pessimistic situation on the market occurs. From the table 2 it is also clear that ranking function solution is dominated by bound and decomposition method solution as its profits are less in both pessimistic and optimistic situation and the average profit value is the same in bound and decomposition method.

We also have to point out that in the presented case study the three solutions for bound and decomposition method are the same, and in this particular situation we know which shares should be selected, but that will not always be the case.

9 Conclusion

This paper offers a possibility to solve portfolio optimization problem with uncertain share profit values using two fuzzy approaches – bound and decomposition method and ranking function method. The results are compares with *p*-fractile approach. The problem takes into account the fact that profits of the shares could not be the same as predicted, but real value of share profit could be different. This paper concentrates on two fuzzy approaches for portfolio optimization model, which use defuzzification method to solve the problem. The presented methods are described on a case study, provided with computation experiments and comparison of results for all presented approaches.

Acknowledgements

This research was supported by the Czech Science Foundation, grant No. P403/12/1947 and by the project F4/11/2013 founded by the Internal Grant Agency of the University of Economics Prague.

References:

[1] T. Allahviranloo, F. Hosseinzadeh Lotfi, M. Kh. Kiasary,N. A. Kiani, L. Alizadeh Solving Fully Fuzzy Linear Programming Problem by the Ranking Function. In: *Applied Mathematical Sciences*, **vol.2**, no.1 (2008), pp. 19–32.

[2] G. Bortolan, R. Degani, A review of some methods for ranking fuzzy subsets, *Fuzzy Sets and Systems* **15** (1985) 1-19.

[3] C. H. Cheng, A new approach for ranking fuzzy numbers by distance method, *Fuzzy Sets and Systems* **95** (1998) 307-317.

[4] M. Inuiguchia, J. Ramík, Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem, *Fuzzy Sets and Systems* **111** (2000) 3-28.

[5] M. Jayalakshmi, P. Pandian A New Method for Finding an Optimal Fuzzy Solution For Fully Fuzzy Linear Programming Problems. In: *International Journal of Engineering Research and Applications (IJERA)* **vol. 2**, Issue 4 (2012), pp. 247-254.

Sustainable Performance Measurement Using Data Envelopment Analysis

Alena Kocmanová[1], Marie Pavláková Dočekalová[2]

Abstract. The aim of this paper is to analyse the sustainable corporate performance (SCP) in Czech manufacturing industry through quantitative efficiency analysis - Data Envelopment Analysis (DEA). SCP is defined by environmental, social, corporate governance (ESG) and economic indicators. Utilization of DEA method for evaluating SCP, seems to be appropriate, especially because it does not evaluate only one indicator, but a set of different indicators that determine the level of efficiency achieved by a company. For the calculation of the efficiency of companies in the framework of DEA methods were used classical method CCR and BCC model, both output oriented. Inputs and outputs are the key ESG and economic performance indicators of evaluation system for each company within the meaning of effective/ineffective SCP. The combination of DEA and SCP is a very effective tool for assessing the efficiency and the identification of best practices leading to sustainability of companies.

Keywords: data envelopment analysis, sustainable corporate performance, benchmarking.

JEL Classification: C67
AMS Classification: 90C05

1 Introduction

Sustainability and the associated corporate performance have for long been in the focus of foreign companies, and the same trend is nowadays beginning to assert itself also among Czech companies. Notwithstanding the fact that there is no single definition or interpretation of the concept of sustainable corporate performance (SCP), sustainability nevertheless is becoming one of important measures for the evaluation of companies and their success in a broader comparison. The underlying reason for the introduction of non-financial indicators into corporate governance is the fact that financial indicators very often do not suffice for the assessment of corporate performance. Due to the influence of global markets and global trends, it is necessary to evaluate the corporate performance in a comprehensive manner using sustainability indicators, integrating strategy into a unified Integrated Reporting [4], [2], [7]. In view of such trends, it is important that both financial and non-financial corporate performance indicators be taken into consideration and thus exploited for the enhancement of the companies' sustainable development. The research conducted in 2011-2013 and supported by a grant project of the Czech Republic entitled "Construction of Methods for Multifactor Assessment the Performance of a Company Complex Performance in Selected Sector" dealt with the drafting of environmental, social and corporate governance (ESG) as well as economic indicators for companies from the process industry according to CZ NACE (section C, for parts 10 to 29). The results of the research and the subsequent verification in practice indicate that the ESG and economic performance indicators designed in this manner are relevant for Czech companies operating in the process industry [18], [19], [20], [21], [6]. The aim of this paper is increase understanding of the uses of ESG and economic performance indicators to evaluate the effectiveness of companies from the process industries, and thus also the SCP using the Data Envelopment Analysis (DEA) as a useful benchmarking tool. According to Zhu, 2003, the use of the DEA as a benchmarking tool is a process of defining valid measures for comparing the performance of decision-making units of the reference set, which is used to determine the relative positions of companies comprising the reference set, which thus generates a generating a performance standard (quality standard). In this way, DEA can be considered a benchmarking tool, because the production frontier designed represents an empirical standard of performance.

[1] Brno University of Technology, Department of Economics, Kolejní 4, 613 00 Brno, Czech Republic, e-mail: kocmanova@fbm.vutbr.cz.
[2] Brno University of Technology, Department of Economics, Kolejní 4, 613 00 Brno, Czech Republic, e-mail: docekalova@fbm.vutbr.cz.

2 Theoretical overview

The concept of sustainability has its roots in the broader concept of sustainable development. Sustainable development is defined as an economic development that meets the needs of the present without compromising the ability of future generations to meet their own needs [3]. The concept of a sustainable company, therefore, emphasizes the harmonious and balanced development of environmental, social and economic pillars [8]. The definition of sustainability is related to the concept of strategy known as the sustainable development strategy in relation to the company according to [12]. Significant changes in corporate strategy and management relating to sustainable thinking are associated with the late 20th century. Some authors such as [22] point out that sustainability requires a transformation of thinking and commitment on the part of the management as well as key stakeholders. Other authors [16], [26], [10] who studied sustainability consider it a management framework that integrates environmental and social performance to the economic performance of the company. As reported by [9], [16], [26] the measuring of sustainability should include key indicators based on economic, environmental and social relationships. But sustainability is not possible without effective corporate governance in the company. Therefore, efforts are being made to add corporate governance aspects to environmental, social and economic aspects.

2.1 Sustainable corporate performance

Sustainable corporate performance can be defined as the integration of not only ESG performance but also of economic performance, see Tab. 1.

	Performance	**Indicators**	
Sustainable corporate performance (SCP)	Environmental	ESG	Measures (units)
	Social		
	Corporate Governance		
	Economic	Economic	

Table 1 The general concept of sustainable corporate performance

A pre-requisite for the measurement of environmental, social and economic performance of a company are key performance indicators (KPI). In order to adequately capture the relationship between environmental, social, corporate governance (ESG) and economic performance and progress in implementing sustainability strategies, it is necessary to develop and use appropriate financial and non-financial performance indicators. Measuring SCP is associated with the introduction of financial and non-financial indicators in corporate governance. The most common classification of indicators seems to divide the approach to the performance assessment using financial and non-financial indicators, as many Czech and foreign authors agree [14], [11], [5]. It is important to create measurable and relevant objectives of sustainability and appropriate key performance indicators.

2.2 Benchmarking

Benchmarking is an excellent way of seeking and achieving success through an organic growth, i.e. growth based on one's own performance" [17]. For purposes of doing performance benchmarking (metric benchmarking) the DEA and the Stochastic Frontier Analysis (SFA) is a usable method. In this article, financial and non-financial ESG and economic performance indicators are used as the basis of performance benchmarking. Mathematical and statistical methods, especially the PCA method of factor analysis, were used to determine the ESG and economic performance indicators. Factor analysis is based on the model of factor analysis [23]:

$$Y_j = \sum_{k=1}^{m} w_{jk} F_k + e_j, \; j = 1,..., p \tag{1}$$

where

w_{jk} factor loading,
F_k k-th common factor,
e_j is specific part of the variable j.

The PCA method of factor analysis resulted in establishing a conceptual framework of financial and non-financial ESG and economic performance indicators for companies in manufacturing industry CZ-NACE.

3 Materials and methods

3.1 Data and variables

Calculations were performed on large industrial companies from CZ- NACE 20 - Manufacture of chemicals and chemical products, 22 - Manufacture of rubber and plastic products, 24 - Manufacture of basic metals and 25 - Manufacture of fabricated metal products, except machinery and equipment. These industries were chosen because process industry companies exert a significant impact on both the environment and the society. 50 % of companies have legal form stock company and 50 % are Ltd. 16 % companies have less than 250 employees, 55 % have 250 – 500 employess, 10 % 501 – 750 employees and 19 % more than 750 employees in 2012. To determine the ESG and economic performance indicators, statistical methods, particularly PCA of factor analysis, were used. For a detailed description of the construction of indicators [18], [19], [20], [21], [6].

- Selected environment variables: EN1: Total air emissions (t) - include both direct and indirect emissions; and EN2-Total annual waste production (t).
- Social performance indicators: LA1: Turnover rate (%) – expresses the ratio of terminated employment contracts to the average number of employees and LA2: Number of employees.
- CG1: Corporate Governance remuneration indicator, i.e. the total financial amount of remuneration to the Board of Directors and the Supervisory Board (in thousands CZK), was selected from the Corporate Governance field.
- Economic performance indicators: EC1: Added Value (in thousands CZK) and EC2: Cash Flow (in thousands CZK).

The first step of data processing was a quality check carried out in order to find out whether there were erroneous or missing data. The following statistic measures were computed in order to get the basic knowledge of performance indicators: mean, range, standard deviation and variance. To understand the relationship between the variables, a correlation analysis was performed. Correlation coefficients between the input variables and the output variables, i.e. the Added Value and Cash Flow, are listed in Tab. 2.

Indicator	Added Value				Cash Flow			
	2008	2009	2010	2011	2008	2009	2010	2011
EN1	0.189	0.145	0.125	0.272	0.218	0.066	0.023	0.242
EN2	0.835**	0.829**	0.846**	0.900**	0.867**	0.339	0.548**	0.857**
LA1	-0.382	-0.404	-0.346	-0.277	0.0,373	-0.446*	-0.345	-0.340
LA2	0.833**	0.881**	0.933**	0.876**	0.860**	0.446*	0.692**	0.848**
CG1	0.769**	0.563**	0.317	0.541**	0.703**	0.169	-0.080	0.449*

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Table 2 Correlation coefficients

3.2 DEA models

DEA is widely used in economics: it is, e.g., used for evaluating the effectiveness of R&D investments, assessing the level of competitiveness of countries [27] performance evaluation and benchmarking etc . DEA approaches have also been widely used to assess the relative effectiveness of research projects in industry, agriculture, banking institutions, telecommunication companies, and for university rankings [25], [24]. Two models are computed – an input oriented model with constant returns to scale (CCR model) [13] and an input oriented model with variable returns to scale (BCC model) [1]. The proposed model includes 22 companies of manufacturing industry. Each of these companies is considered as a homogeneous production unit (decision-making unit, DMU), $k = 12$. In this model, there are five inputs – $i = 5$ (*Total air emissions, Total annual waste production, Turnover rate, Number of employees* and *Corporate Governance remuneration indicator*) and two outputs – $j = 2$ (*Added Value* and *Cash Flow*). DMU_k consumes X_{ik} of the input i and produces Y_{jk} quantity of the output j. *DMUq* consumes x_{jq} of the input i and produces y_{iq} quantity of the output j. u_i and v_j are individual weights assigned so as to maximise the efficiency and ε is infinitesimal constant which ensures that all weights of inputs and outputs will be positive.

Mathematical model of primary input oriented CCR model [15]:

$$\max z = \sum_{i}^{5} u_i y_{iq}, \tag{2}$$

on conditions:

$$\sum_{i}^{5} u_i y_{iq} \leq \sum_{j}^{2} v_j x_{jk}, \tag{3}$$

$$\sum_{j}^{2} v_j x_{jq} = 1, \tag{4}$$

$$u_i \geq \varepsilon\ , \tag{5}$$

$$v_j \geq \varepsilon\ . \tag{6}$$

Mathematical model of primary input oriented BCC model [15]:

$$\max z = \sum_{i}^{5} u_i y_{iq} + \mu, \tag{7}$$

on conditions:

$$\sum_{i}^{5} u_i y_{iq} + \mu \leq \sum_{j}^{2} v_j x_{jk}, \tag{8}$$

$$\sum_{j}^{2} v_j x_{jq} = 1, \tag{9}$$

$$u_i \geq \varepsilon, \tag{10}$$

$$v_j \geq \varepsilon, \tag{11}$$

μ - arbitrary. (12)

The difference compared to the CCR model is the introduction of the convexity condition. μ is a dual variable associated with the convexity condition:

$$e^T \lambda = 1, \tag{13}$$

This additional constraint gives the frontiers piecewise linear and concave characteristics. The coefficient of technical efficiency *z* derived from the DEA model is relative, expresses efficiency of the DMU within the studied group of DMUs. If equal to one, it means that in the group there is no more effective unit. If the value of *z* is less than 1, there is at least one more DMU which is more effective.

4 Results and discussion

In the DEA-model companies are benchmarked against the efficient frontier, which is constructed by efficient companies. Calculation results according to the CCR and the BCC models were averaged for the 2008 – 2011 monitoring period, and the order according to individual models was established. It follows from a comparison between the CCR and the BCC models that the model with variable returns to scale will label a greater number of enterprises as effective.

Company	2008		2009		2010		2011		Average		Ranking	
	CCR	BCC	CCR	BCC	CCR	BCC	CCR	BCC	CCR	BCC	CCR	BCC
A	0.37	0.43	0.36	0.43	0.45	0.47	0.40	0.47	0.40	0.45	19	22
B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1
C	0.94	1.00	0.53	1.00	0.38	0.63	0.45	1.00	0.58	0.91	17	18
D	0.37	0.70	0.50	0.67	0.35	0.63	0.28	0.74	0.38	0.69	20	20
E	0.53	1.00	0.61	1.00	1.00	1.00	1.00	1.00	0.79	1.00	12	1
F	1.00	1.00	0.95	0.97	1.00	1.00	1.00	1.00	0.99	0.99	8	14
G	0.43	1.00	0.39	1.00	1.00	1.00	0.87	1.00	0.67	1.00	16	1
H	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1
I	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1
J	0.59	1.00	0.73	1.00	0.94	1.00	0.93	1.00	0.80	1.00	11	1
K	1.00	1.00	0.77	1.00	0.59	1.00	0.66	1.00	0.76	1.00	14	1
L	1.00	1.00	0.49	1.00	0.33	0.74	0.45	0.83	0.57	0.89	18	19
M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1
N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1
O	0.11	1.00	0.15	1.00	0.21	1.00	0.24	1.00	0.18	1.00	22	1
P	0.21	0.61	0.32	0.61	0.23	0.60	0.24	0.66	0.25	0.62	21	21
Q	0.56	1.00	0.79	1.00	0.72	1.00	1.00	1.00	0.77	1.00	13	1
R	1.00	1.00	1.00	1.00	1.00	1.00	0.72	0.95	0.93	0.99	10	15
S	0.91	0.99	0.84	0.86	1.00	1.00	1.00	1.00	0.94	0.96	9	17
T	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1
U	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1
V	0.56	1.00	0.63	0.95	0.80	0.99	0.92	1.00	0.73	0.99	15	16

Table 4 Efficiency scores of companies

The DEA method potential consists in identifying the best practice in the industry. As an example, take the “D" Company. Its efficiency score according to the CCR model was 0.2835 (28.35 %) in 2011. The target input levels for the "D" company are defined as the weighted average of inputs of relevant peer units identified in the dual model. The inefficient "D" Company should behave as a combination of peer units "F", "I" and "T". Table 5 gives the benchmarks for the input variables that will make sure that the company becomes effective at the current level of Added Value and Cash Flow.

Indicators	Actual state	Benchmark	Potential improvement (%)
Total air emissions (t)	1.80	0.51	-71.65
Total annual waste production (t)	6543	48.28	-95.07
Turnover rate (%)	32.52	0.14	-99.26
Number of employees	206.00	58.41	-99.56
CG remuneration (th. CZK)	35.00	9.92	-39.26

Table 5 Potential improvements of inputs - “D” company

5 Conclusions

In the paper, DEA application using financial and non-financial ESG criteria to sustainable corporate performance was studied. A framework for the setting up of benchmarks and the identification of the best practice in the industry was defined, which would ensure that inefficient companies might become efficient, i.e. achieve a sustainable corporate performance. Efficient companies are referred to as "peer companies" and determine the efficiency frontier, i.e. they use the minimum quantity of inputs to produce the same quantity of outputs. The advantage of using DEA is not only that it can identify gaps in corporate performance, but also that it makes it possible to quantify those performance gaps - such information then becomes important source material for quality decisions by the management.

Acknowledgements

This paper is supported by the grant No. 14-23079S Measuring Corporate Sustainability in Selected Sectors of The Czech Science Foundation.

References

[1] Banker, R.D., Charnes, R.F., and Cooper W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* **30** (1984), 1078–1092.
[2] Bell, S., and Morse, S.: *Sustainability Indicators: Measuring the Immeasurable.* Earths can publishes in association with the International Institute for Environment and Development, New York, USA, 2008.
[3] Brundtlandová, G.H.: *Naše společná budoucnost.* Zpráva Světové komise pro životní prostředí a rozvoj. Praha, Academia, 1991.
[4] Cokins, G.: *Performance management: integrating strategy execution, methodologies, risk, and analystics.* John Wiley & Sons Inc., New York, 2009
[5] Dixon, J.R, Nanni, A.J, and Vollmann. T.E.: *The new perfomance challenge: measuring operations for world-class companies.* Homewood, IL, Dow Jones-Irwin, 1990.
[6] Dočekalová, M.: Construction of corporate social performance indicators for czech manufacturing industry. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **2** (2013), 309 □ 315.
[7] Eccles, R.G, and Krzus, M.P.: *One report.* John Wiley & Sons Inc., New York, 2010.
[8] Elkington, J.: *Cannibals with forks: The triple bottom line of 21st century business.* New Society Publishers, 1998.
[9] Epstein, M. J.: *Making sustainability work: best practices in managing and measuring social and environmental impacts.* Sheffield, Greenleaf, 2008.
[10] Epstein, M. J., and Roy, M. J.: Improving sustainability performance: specifying, implementing and measuring key principles. *Journal of General Management* **1** (2003), 15–31.
[11] Grigoroudis, E., Orfanoudaki, E. and Zopounidis, C.: Strategic performance measurement in a healthcare organization: A multiple criter ia approach based on balanced score. *Omega* **40** (2012), 104–119.
[12] Harts, S.: A natural ressource-based view of the firm. *Academy of Management Review* 4(1995), 966 □ 1014.
[13] Charnes, A., Cooper, E., and Rhodes, E.: Measuring the efficiency of decision- making units, *European Journal of Operational Research* **2** (1978), 429 □ 444.
[14] Ittner, C, Larcker,D, and Randall, T.: Performance Implications of Strategic Performance Measurement in Financial Services Firms. *Accounting, Organizations & Society* **28** (2003), 715 □ 741.
[15] Jablonský, J.: *Operační výzkum: kvantitativní modely pro ekonomické rozhodování*. Praha, Professional Publishing, 2002.
[16] Johnson, M.: *Stakeholder dialogue for sustainable service*. Karlstad, Karlstad University Studies, 2007.
[17] Karlöf, B., and Östblom, S.: *Benchmarking: Jak napodobit úspěšné*. Praha, Victoria Publishing, 1995.
[18] Kocmanová, A., and Dočekalová, M. Construction of the economic indicators of performance in relation to environmental, social and corporate governance (ESG) factors. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **4** (2012), 195–205.
[19] Kocmanová, A., and Dočekalová, M. Corporate sustainability: environmental, social and corporate performance. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **7** (2011), 203–208.
[20] Kocmanová, A., Dohnal, M., Meluzín, T.: Qualitative simple equation less models as simple integrators of vague sustainability knowledge items. *Transformations in Business& Economics* **3** (2011a), 187 □ 196.
[21] Kocmanová, A., and Šimberová, I.: Modelling of corporate governance performance indicators. *Engineering Economics* **5** (2012), 485 □ 495.
[22] Laszlo, Ch.: *The sustainable company: How to create lasting value through social and environmental performance*. Washington, Island Press, 2003.
[23] Meloun, M., and Militký, J.: *Statistická analýza experimentálních dat*. Praha, Academia, 2004.
[24] Ntantos P.N , and Karpouzos D.K. Application of Data Envelopment Analysis and Performance Indicators to Irrigation Systems in Thessaloniki Plain (Greece). *International Journal of Engineering and Natural Sciences* **4 (**2010), 51 □ 57.
[25] Sarkis, J., and Weinrach, J.: Using data envelopment analysis to evaluate environmentally conscious waste treatment technology. *Journal of Cleaner Production* **9** (2001), 417 □ 427.
[26] Schaltegger, S., and Wagner, M.: *Managing sustainability performance measurement and reporting in an integrated manner, sustability accounting as the link between the sustainability balanced scorcared and sustainability reporting*. Dordrecht, Springer, 2006.
[27] Staníčková, M., and Malecký, L.: Hodnocení konkurenceschopnosti visegrádské čtyřky prostřednictvím aplikace CCR vstupově orientovaného modelu analýzy obalu dat, Scientific Papers of the University of Pardubice **22** (2011), 176–188.

Evaluation of the effects of inflation targeting policy by a modified VAR model

Jan Kodera[1], Quang Van Tran[2]

Abstract. In our contribution we examine the relations between economic variables relevant for inflation targeting. For this purpose both directly observable variables as well as a set of unobservable factors are used. These unobservable factors are characterized by additional time series of auxiliary economic observable variables bearing the influence (dynamic) of these factors. To perfectly grasp the influence of factors usually the bigger amount of auxiliary variables are required. For this experiment we use macroeconomic data on inflation (CPI), interest rate, and the GDP growth rate. To generate the factors, we take data on technological level and monetary policy. For a proper estimation of parameters in the VAR model, we use a two step procedure. In the first step we apply the principal component method to estimate the factors from auxiliary time series. In the second step a standard method of estimation technique for a VAR model is used, where the unobservable factors are represented by their estimates from the first stage. The results of a modified VAR model obtained by using this approach on the data of the Czech economy are compared with the ones from traditional methods and they confirm the superiority of this method to traditional ones.

Keywords: Monetary policy, Inflation targeting, Vector autoregression, Factor analysis

JEL classification: C51
AMS classification: 91G70

1 Introduction

When targeting inflation, a central bank often uses its short-term interest rate to manage the main economic variables as the price level and economic growth. This is achieved through the so called transmission mechanism of monetary policy discussed in the literature [10],[3]. The effect of monetary policy on real economy is usually quantified either by DSGE model approach or VAR model approach. With the technique using VAR model, only a limited number of variables can be included as we need a positive degree of freedom for the model. This causes two main problems. Firstly by doing so, the impact of a lot of other variables that might affect the main economic variables of our interest is omitted. Secondly, the impulse response analysis of VAR model can be proceeded with those variables present in the model. To alleviate these disadvantages, Bernanle et al [1] propose a technique which enable to incorporate into a traditional VAR model a vast number of other variables in the form of common factors. The model is therefore named as factor augmented VAR model (FAVAR). Further we use Granger causality test to evaluate the mutual impact of or main economic variable and factors. To our knowledge, such a relationship has not been investigated so far in the Czech Republic (the only exception is the work of Brůna and Brada [4], who investigate the sensitivity of short term interbank interest rates PRIBOR to changes in repo rate Czech National Bank). To fill this gap in the literature, in this experiment, we use the FAVAR approach to examine the effect of inflation targeting monetary policy of the Czech National Bank (CNB) from period 2000 - 2013. Our objective is to quantitatively evaluate the impact of interest rate, technological level and monetary quantities on inflation level and economic growth in the Czech economy in the aforementioned period.

[1]University of Economics, Prague,Department of Statistics and Probability, Winston Churchill Sq. 4, Prague, Czech Republic, kodera@vse.cz
[2]University of Economics, Prague,Department of Statistics and Probability, Winston Churchill Sq. 4, Prague, Czech Republic, tran@se.cz

2 Factor Augmented VAR model

In a traditional VAR model estimation, a vector X of size $m \times 1$ of observable and stationary variables is required. The standard estimation technique allows us to first estimate a VAR model in the so called reduced form. Under certain assumptions, a structural form of the model can be retracted from the reduced form and the impulse response analysis can be performed. With respect to the length T of dataset available to a researcher, the dimension of vector X must be relatively small to ensure the positiveness of the degree of freedom. This is important for the calculation of all statistics indicating the significance of the model.

Nevertheless, economic variables included in X can be affected by other determinants absent in X. Therefore, central banks when conducting monetary policy, they consider a wide range of economic indicators. To take into account the impact of these variables on main economic quantities as well as to avoid the omission of some important variables problem, Bernanke et al [1] propose to extend the model by some unobservable factors F. As the number of factors is reasonably small, the size of factor vector is $n \times 1$, the positiveness of the degree of freedom of a model is still conserved while the effect of many other economic quantities is included into the model. The VAR model augmented by these factors (FAVAR) is specified as follows

$$\begin{bmatrix} X \\ F \end{bmatrix}' = \Phi(L) \begin{bmatrix} X_{-1} \\ F_{-1} \end{bmatrix}' + \nu, \tag{1}$$

where $\Phi(L)$ is a lag polynomial of some order and ν is a vector of error terms.

The unobservable factors model (1) are estimated by principal components analysis technique. This technique is numerically robust and computationally efficient, and is econometrically consistent for the latent factors F under the standard assumptions discussed in Stock and Watson [11]. As these factors represent the effect of a variety of economic time series on the main economic economic variables in the model, they are extracted from a vector Z of an additional informational time series of size $p \times 1$ ($p \gg m + n$). Since the standard PCA analysis imposes the restriction that the covariance matrix of F has to be be the identity matrix, an unobservable factor can be extracted as

$$\hat{F} = \sqrt{T}\hat{V}, \tag{2}$$

where $\hat{V}$ is an eigenvector of matrix $Z'Z$ corresponding to largest eigenvalue. If we want to introduce to the model n factors, n eigenvectors corresponding to n largest eigenvalues are used.

3 The inputs for modeling the development of an economy

Inflation targeting regime is the last format of monetary policy conduct. The main objective of monetary policy in this regime is the price stability[1]. Other objectives in this regime as to support economic growth and high employment are also possible, but they must be subordinate to the primary price stability objective. When targeting inflation, first a central bank has to set a target level of inflation. Then she often uses its main monetary policy tool as the short term interest rate to steer the actual inflation level in the economy toward its target. If the central bank succeeds in managing the inflation level in the economy, all agents would take its inflation target into their expectation about the future level of inflation. It is very important as expectation always plays a key role in any economy.

Focusing on inflation as the only primary goal of monetary policy does not imply that implementing monetary policy in the inflation targeting regime is less complicated than in other formats. Conversely, it requires the monetary authority to fully understand how the economy works. As inflation is a result of many interactions among many economic variables and their relationships, central banks often collect and use economic data from a great deal of areas. For example, in its quarterly inflation report the Czech National Bank, which has been targeting inflation since 1998, states its inflation assessment on the development of inflation in the economy is based on the economic indicators from seven areas. They are:

- Inflation
- Import prices and producer prices

[1] The price stability does not necessarily mean that there has to be a zero inflation level in the economy. On the contrary, inflation target in most developed economies is set at level of 2%, i.e. relatively low and stable rate of inflation.

- Demand and output
- The labor market
- Financial and money development
- Balance of payments
- The external environment.

According to the report, the Czech central bank uses up to 60 individual economic indicators to evaluate the development of the Czech economy.

Today, the behavior of an economy can be model either by DSGE approach or the VAR approach. Though the DSGE approach is theoretically better reasoned due to its microeconomic backgrounds, the results of this modeling approach are not superior to the ones obtained by using the VAR approach. Therefore, the VAR approach, particularly its modifications, is still very popular. In this our experiment, the FAVAR modification is used. Besides the three main observable variables, we also construct an informational vector, whose elements are chosen from the list of indicators used by the Czech Central Bank. Our choice is guided by the emphasis of the role of technology and monetary policy indicators. As the transmission mechanism has not been fully understood yet [10],[3][2], the impact of monetary policy interest rate changes on real economy is difficult to be evaluated. For this reason, instead of modeling relationship among the monetary policy interest rate (repo rate) with inflation and economic growth, we replace it by using long term interest rate represented by the return of average long term government bond. As the Czech economy is a small open economy, exchange rate regime plays very important role in transmitting monetary policy into real economy [12],[5], nominal and real exchange rates are also part of the informational vector.

4 Data and their econometric analysis

To implement our objective, as the observable economic quantities we select these time series: from the database of the Czech statistical office series from period 1st quarter of 2000 to 4rd quarter of 2013 on quarterly real Gross Domestic Product (GDP), and Consumer Price Index (CPI), from the CNB database Yield on the basket of government bonds with an average residual maturity of 10 years. As informational time series for industry and technology the following series are chosen: Industrial Production Index (IPI), Sales from Industrial Activity (SIA), New Industrial Order (NIO), Employment in Industry (EI), Wages in Industry (WI) and Unemployment Rate (UR). As the informational time series for monetary policy from the database of the Czech National Bank these series are selected: monetary aggregates M1, M2 and M3, nominal and real effective exchange rates (NEER, REER), interbank short term interest rate for three months (PRIBOR 3M) and interest rates of commercial banks on CZK loans provided to Czech nonfinancial firms (IRNFF).

To ensure the stationarity of all series used in our study, we transform all series as follows. Series of interest rates are transformed by simple differencing. The rest of them are detrended using Hodrick Prescott filter with lambda = 6400. Then the transformed series are tested for stationarity using Augmented Dickey Fuller unit root test. The results confirm that all are stationary after transformation. After that from informational series two factor named F1 and F2 series are generated according to formula (2).

Series IR, GDP and π together with two unobservable factor series F1 and F2 are used to estimate a pair-wise reduced VAR models. The maximum length of lags we use is 2 as we consider the period of six quarters is long enough for a monetary policy measure to induce its impact on real economy. The estimation results are reported in Table 1. Together with the estimates in each lag which are the values in first line, we also report the value of standard errors of the estimates which are the numbers in the second line in parentheses and finally the numbers in the third line in squared brackets ares values of t-statistic.

The results of our econometric analysis shows that in the case of the Czech Republic interest rate affects economic quantities with time delay, the sing of its effect is consistent with economic theory [8].

[2] this phenomenon does not indicate that monetary policy has become less effective [9].

Table 1 The results of FAVAR model estimation

	IR	GDP	π	F1	F2
IR(-1)	0.490305	-0.227282	0.100804	-0.202432	0.172512
	(0.18063)	(0.66223)	(0.50862)	(0.23522)	(0.33212)
	[2.71442]	[-0.34321]	[0.19819]	[-0.86062]	[0.51942]
IR(-2)	-0.025782	-2.348300	-0.811105	1.079296	0.022150
	(0.18858)	(0.69137)	(0.53100)	(0.24556)	(0.34673)
	[-0.13672]	[-3.39661]	[-1.52752]	[4.39516]	[0.06388]
GDP(-1)	-0.176706	0.326108	-0.105774	0.112070	0.028770
	(0.08362)	(0.30658)	(0.23546)	(0.10889)	(0.15375)
	[-2.11317]	[1.06371]	[-0.44922]	[1.02919]	[0.18712]
GDP(-2)	0.146292	-0.608292	-0.146709	0.146300	-0.077786
	(0.07183)	(0.26336)	(0.20227)	(0.09354)	(0.13208)
	[2.03652]	[-2.30972]	[-0.72531]	[1.56399]	[-0.58892]
$\pi(-1)$	0.066172	-0.017459	0.960097	0.053440	0.107596
	(0.07063)	(0.25895)	(0.19889)	(0.09198)	(0.12987)
	[0.93686]	[-0.06742]	[4.82734]	[0.58101]	[0.82848]
$\pi(-2)$	-0.064826	-0.010165	0.016526	-0.051009	-0.108972
	(0.07029)	(0.25772)	(0.19794)	(0.09154)	(0.12925)
	[-0.92221]	[-0.03944]	[0.08349]	[-0.55724]	[-0.84311]
F1(-1)	-0.373041	-1.017052	-0.434142	1.037277	0.021948
	(0.20114)	(0.73741)	(0.56636)	(0.26192)	(0.36983)
	[-1.85467]	[-1.37922]	[-0.76655]	[3.96030]	[0.05935]
F1(-2)	0.247520	-0.835048	-0.209312	0.036703	-0.432201
	(0.18560)	(0.68044)	(0.52260)	(0.24168)	(0.34125)
	[1.33365]	[-1.22723]	[-0.40052]	[0.15187]	[-1.26651]
F2(-1)	-0.041894	0.557817	-0.082217	-0.102226	0.753080
	(0.11568)	(0.42411)	(0.32573)	(0.15064)	(0.21270)
	[-0.36215]	[1.31527]	[-0.25241]	[-0.67862]	[3.54060]
F2(-2)	0.169248	-0.019369	0.580932	0.089029	-0.143426
	(0.11914)	(0.43681)	(0.33549)	(0.15515)	(0.21907)
	[1.42054]	[-0.04434]	[1.73162]	[0.57383]	[-0.65471]
C	-0.212870	3.111925	3.210807	-0.298781	0.063629
	(0.89811)	(3.29270)	(2.52892)	(1.16952)	(1.65136)
	[-0.23702]	[0.94510]	[1.26964]	[-0.25547]	[0.03853]

Table 2 Granger causality test results

Null hypothesis	No of Obs	stat	p-value
π does not Granger Cause IR	46	0.93820	0.3996
IR does not Granger Cause π		0.94037	0.3987
GDP does not Granger Cause IR	46	1.22590	0.3040
IR does not Granger Cause GDP		0.45760	0.6360
F2 does not Granger Cause IR	46	1.48061	0.2394
IR does not Granger Cause F2		1.06363	0.3545
F1 does not Granger Cause IR	46	0.27630	0.7600
IR does not Granger Cause F1		1.21578	0.3069
GDP does not Granger Cause π	46	0.20575	0.8149
π does not Granger Cause GDP		3.76400	0.0316
F2 does not Granger Cause π	46	3.05511	0.0580
π does not Granger Cause F2		1.03400	0.3647
F1 does not Granger Cause π	46	0.48647	0.6183
π does not Granger Cause F1		6.12197	0.0047
F2 does not Granger Cause GDP	46	4.24318	0.0211
GDP does not Granger Cause F2		2.11026	0.1342
F1 does not Granger Cause GDP	46	3.48110	0.0401
GDP does not Granger Cause F1		0.85762	0.4316
F1 does not Granger Cause F2	46	5.45816	0.0079
F2 does not Granger Cause F1		4.71680	0.0143

While its effect on inflation is not so strong and statistically less significant, the effect of interest rate on economic growth is obvious both in terms of its size and its statistic relevance. This finding is consistent with theory as well as the finding of other authors [2],[6],[13]. As far as the impact of those unobservable factors is concerned, we detect only a statistically weaker effect of the second factor on inflation. In our opinion, this result can be a consequence of grouping a variety of economic indicators with different impact in terms of sign on the main economic quantities as inflation and growth. Further, we use Granger causality test to examine the direction of causal relationship between a interest rate change and those variables included in the model. The results of the test are shown in Table 2. The test result does not confirms the impact of interest rate rate change on the main economic variables of our interest in the VAR model. On the other hand, it also shows that unobserable factors may have some impact a on GDP. Other relevant causality effect in the sense suggested by Granger as well as all reverse causal relationships are not observed.

5 Conclusion

In this experiment, we investigate the impact of interest rate change on two important economic variables which is economic growth and inflation in the case of the Czech Republic. We have used a VAR model and Granger causality test on economic data of the Czech economy from period 2000 – 2013. We have expanded our VAR model by two factors extracted from an additional informational matrix consisted of economic indicators basically from sector industry and those related to monetary policy. The results of our experiment with VAR models have clearly shown the link between interest rate and GDP growth and, to some extent, also to inflation. Interest rate affect these two variables with delay and this result is in accordance with theory. While we have not traced the impact of the two factors obtained with principle components analysis from various economic indicator, the results of the Granger direction of causality

has shown they may have direct impact on GDP growth and inflation. At this stage, we do not discuss the magnitude of these links in details and it will be subject of our further research. Another direction of our future research might be to extract different unobservable factors from different groups of indicators in order to suppress their mutually eliminating power.

Acknowledgements

The authors are grateful for the financial support of Czech Science Foundation No. P402/12/G097.

References

[1] Bernanke, B. S., Boivin, J., Eliasz, P. S.: 2005. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics* **120** (2005), 387–422.

[2] Bernanke, B. S., Getler, M.: Inside the Black Box: The Credit Channel of Monetary Policy Transmission, *The Journal of Economic Perspectives* **9** (1995), 27–48.

[3] Boivin, J, Giannoni, M.: Assessing Changes in the Monetary Transmission Mechanism: A VAR Approach. *Federal Reserve Bank of New York Economic Policy Review*, **8** (2002), no. 1, 97–111.

[4] Bruna, K., Brada, J.: An analysis of PRIBOR interest rates sensitivity to changes in Czech national bank repo rate, *Politická ekonomie* **5** (2004), 602–21.

[5] Cevik, S., Teksoz, K.: Lost in Transmission? The Effectiveness of Monetary Policy Transmission Channels in the GCC Countries, *IMF Working Paper* WP/12/191 (2012).

[6] Dedola, L., Lippi, F.: The Monetary Transmission Mechanism: Evidence from the Industries of five OECD Countries, *European Economic Review* **49** (2005), 1543–1569.

[7] Czech National Bank: *Inflation Report* **1**/2014.

[8] Kodera, J., *Měnová analýza*, 2. upravené vydání, ASPI, Praha, 2007.

[9] Kuttner, K. N., Mosser, P. C.: The Monetary Transmission Mechanism: Some Answers and Further Questions. *Federal Reserve Bank of New York Economic Policy Review* **8** No. 1 (2002), 15–26.

[10] Mishkin, F. S.: The Channels of Monetary Transmission: Lessons for Monetary Policy, *Banque De France Bulletin Digest* **27** (1996), 33–44.

[11] Stock, J.,Watson, M.: Forecasting Inflation, *Journal of Monetary Economics* **44** (1999), 293–335.

[12] Taylor, J.: The Monetary Transmission Mechanism: An Empirical Framework, *Journal of Economic Perspectives* **9** (1995), 11-–26.

[13] Vizek, M.: Econometric Analysis of Monetary Transmission Channels in Croatia, *Croatian economic survey* **10** (2007), 11–44.

Heuristic Approach to the Zone Partitioning Problem in Counting Zones Tariff System

Michal Kohání[1]

Abstract. When designing a tariff system in regional public transportation, there are several approaches how to design it. One of various approaches is the zone tariff where the whole region is divided into the smaller sub-regions - tariff zones. The price of travelling in such tariff depends on the number of travelled zones. Designing of zone tariff system is connected with the decisions about optimal price for travelling and optimal zone partitioning of the region. In this paper we propose a heuristic method for solving of zone partitioning problem in counting zones tariff system. Previously proposed exact method has some drawbacks, especially when solving the problem on large datasets where computational time of exact method increase significantly with the size of the problem. To test behavior of proposed heuristic method we make a computational study on test data from selected region and compare the results of obtained by this method with the solution of exact method in terms of computational time and preciseness of the solution.

Keywords: tariff planning, tariff zones design, IP solver, heuristic method.

JEL Classification: C44, C61
AMS Classification: 90C08, 90C59

1 Introduction

When the transport authorities plan the regional public transportation, one of the problems they deal with is the problem of the tariff and the ticket prices. As was mentioned in [3] and [8], there are various tariff types, such as *distance tariff*, *unit tariff* and *zone tariff*. In the distance tariff the price for travelling depends on the real length of the trip, in the unit tariff system the price is constant for all trips and is independent on the distance. In the zone tariff system the region is divided into smaller sub-regions - tariff zones and the price for travelling depends on the origin and the destination zone and usually on the number of travelled zones on the trip.

The zone tariff systems can be divided into two groups. In the *zone tariff with arbitrary prices*, the prices depend on the origin and destination zones and the number of travelled zones is not important, because prices are given for all pairs of zones separately and arbitrary. In t*he counting zone tariff system*, the price of trip is calculated according to the origin zone, the destination zone and the number of travelled zones. For all trips hold that the price for passing the same number of zones must be the same. Example of a counting zone tariff system in Bratislava region in Slovakia is in the Figure 1.

Another important task in the planning the tariff is how to design the zones and to fix the new fares. Hamacher and Schöbel in [3], Schöbel in [8] and Babel and Kellerer in [1] proposed approaches for the zone design problem with arbitrary prices. Hamacher and Schöbel in [3] and Schöbel in [8] mentioned the solving of the counting zones tariff system where the goal is to design the zones such that the new and the old price for most of the trips are as close as possible. They proposed three different objectives based on fair design, models for the fare problem and zone partitioning and three heuristic approaches. A note on fair fare tariff was mentioned also by Paluch in [7]. Another approach was described by Müller, Haase and Klier in [6], where they formulated model and algorithm for revenue maximizing tariff zone planning.

This paper will be organizes as follows. In the section 2, we present the model of the zone partitioning with given prices and number of zones with the average deviation criterion for counting zones tariff system. In the chapter 3 we describe solution methods for zone partitioning problem and introduce heuristic method based on greedy approach with two different criteria. In the chapter 4 we present numerical experiments on the test network to study computational demands of proposed heuristics.

[1] University of Žilina, Univerzitná 8215/1, Žilina, Slovakia, e-mail: kohani@frdsa.fri.uniza.sk

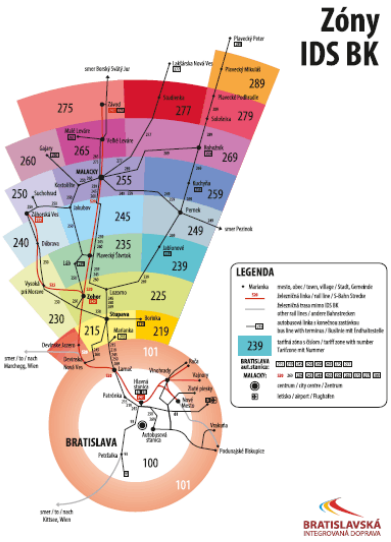


Figure 1: Example of the counting zones tariff system in Bratislava region *(www.bid.sk)*

2 Mathematical model of the zone partitioning with given prices and number of zones

Let all stations in the network of public transport constitute the set of nodes I. The station i and j from set I are connected by the edge $(i,j) \in E$, if there is direct connection by public transport line between these two stations. Symbol E denotes the set of edges. The distance between stations i and j is denoted as d_{ij}. For each pair of stations i and j is c_{ij} the current price of travelling between these two stations. We assume that for each pair of stations i and j $c_{ij}=c_{ji}$.The number of passengers between stations i and j is b_{ij} (OD matrix).

If we want to calculate new price of the trip between nodes i and j in the counting zones tariff system, we need to calculate the number of zones crossed on this trip. The calculation of the number of crossed zones can be easily replaced by the calculation of crossed zone borders as was used in [3] and [6]. We assume that the station can be assigned only to one zone and then the border between zones is on the edge. We will introduce the binary variable w_{rs} for each existing edge $(r, s) \in E$, which is equal to 1 if stations r and s are in different zones and is equal to 0 otherwise. For calculation of the number of crossed borders we need to determine the used path for travelling between stations i and j. We introduce parameter a_{ij}^{rs}, where the used paths will be observed. a_{ij}^{rs} is equal to 1 if the edge (r,s) is used for travelling from station i to station j and 0 otherwise.

When we want to set a new price for travelling in proposed system, there are more possibilities how to do it. Hamacher and Schöbel in [3] and Schöbel in [8] proposed solution of fare problem with fixed zones to obtain new fares for trips with various travelled zones. In [4] and [5] a unit price for travelling per one zone was set. In this paper we use two different unit prices, as was mentioned in [4] – price f_1 for travelling in the first zone and unit price f_2 for travelling in each additional zone. The final new price will be calculated as a sum of the price for the first zone and number of other travelled zones multiplied by the unit price for additional zones. This notion is more natural and often used also in distance tariff, where the average price per kilometer is higher for short trips. New price n_{ij}, determined by the number of crossed zones will be calculated according to this definition as follows (1):

$$n_{ij} = f_1 + \sum_{(r,s)\in V} f_2 a_{ij}^{rs} w_{rs} \tag{1}$$

Construction of the zone partitioning model was inspired by the model of the p-median problem. We introduce binary variables y_i, which represent the "fictional" centre of the zone. Variable y_i is equal to 1 if there is a centre of the zone in node i and 0 otherwise. For each pair of stations i and j we introduce variable z_{ij}. Variable z_{ij} is equal to 1 if the station j is assigned to the zone with centre in the node i and 0 otherwise. We expect to create at most p tariff zones.

When we want to suggest the objective function of the model, there are many possible ways. In [4] they proposed three different objectives based on fair design, in [6] authors formulated the criterion of revenue maximizing. In our model we will use the average deviation between current and new price for all passengers. According to the advices of experts in [8], in this paper we will use the average deviation between current and new price as

a criterion in objective function. The current or fair price between stations i and j will be denoted by c_{ij}. The mathematical model of zone partitioning with fixed prices and number of zones can be written in the form:

$$Minimize\ dev_{avg} = \frac{\sum_{i\in I}\sum_{j\in J}\left|c_{ij}-n_{ij}\right|b_{ij}}{\sum_{i\in I}\sum_{j\in J}b_{ij}} \tag{2}$$

$$subject\,to \sum_{i\in I} z_{ij} = 1,\ for\ j \in I \tag{3}$$

$$z_{ij} \le y_i,\ for\ i, j \in I \tag{4}$$

$$z_{ij} - z_{ik} \le w_{jk},\ for\ i \in I, (j,k) \in E \tag{5}$$

$$\sum_{i\in I} y_i \le p \tag{6}$$

$$z_{ij} \in \{0,1\},\ for\ i,\ j \in I \tag{7}$$

$$y_i \in \{0,1\},\ for\ i \in I \tag{8}$$

$$w_{ij} \in \{0,1\},\ for\ (i,j) \in E \tag{9}$$

Condition (3) ensures that each station will be assigned exactly to one zone. Condition (4) ensures that the station j will be assigned only to the existing centre of the zone. Condition (5) is coupling between variables for allocation of the station to the zone and variables for determining the zone border on the edge (j,k). Condition (6) ensures that we will create at most p tariff zones.

3 Solution methods

As was shown in the model (2) – (9), the objective function (2) in this model is not a linear function. To be able to solve this problem, in [4] and [5] there was proposed the linearization of the model. The linearised model was solved using IP solver Xpress in this case. To determine the optimal values of parameters in the model, a two-phase procedure was used. In the first phase the optimal number of zones was determined. In the second phase *(ZP_ex)*, the model with different settings of parameters f_1 and f_2 and with given number of zones p was solved. The optimal parameter setting was determined as the parameter setting of the solution with the smallest value of the objective function. The major drawback of this process was the time complexity of the problem. Computation time for given parameter p grows rapidly with the increase of the problem size, as was written in [5].

Another approach to solve this problem was proposed in [3] and [8]. In the first stage the optimal price of travelling was calculated. Subsequently, three algorithms were used to calculate the zone partitioning. First algorithm is based on the clustering theory, second algorithm is a greedy algorithm and the last algorithm is based on the spanning tree approach.

Based on ideas of previously mentioned methods we proposed a heuristic method to solve the zone partitioning problem. The goal is to create zones that are approximately the same size. To be able to do it, we apply a greedy approach, where two smallest neighboring zones will be joined to the one new zone. The proposed heuristic method can be described as follows *(ZP_heur)*:

STEP 1: Start with a partition P consisting of $|I|$ zones, each zone contains a single node/station. For each zone Z_i from P calculate the parameter e_i which expresses the size of the zone.

STEP 2: Determine two neighboring zones Z_i and Z_j, where the sum $e_i + e_j$ is minimal.

STEP 3: Join zones Z_i and Z_j to the new zone Z_k and get a new partition P. For new zone Z_k calculate parameter e_k.

STEP 4: If the terminating condition is reached, then stop, else go to STEP 2.

For the zone size calculation we propose two different formulas. The first one calculates the size e_i of the zone i as the average distance of all stops in the zone Z_i to the neighboring zones as follows:

$$e_i = \frac{\sum_{k \in Z_i} \sum_{j \in S_i} d_{kj}}{|Z_i| . |S_i|} \tag{10}$$

where S_i is the set of stops, which are connected with at least one of the stops in the zone Z_i and $|Z_i|$ represents the number of nodes in the zone Z_i. The second approach of calculation the zone size adds the number of inhabitants in the zone as a weight to the average distance as follows:

$$e_i = \frac{\sum_{k \in Z_i} \sum_{j \in S_i} d_{kj} b_k}{|Z_i| . |S_i|} \tag{11}$$

where parameter b_k represents the number of inhabitants in the node k.

As the terminating condition in Step 4 we can use the given number of zones or we can calculate the value of the objective function with different settings of parameters f_1 and f_2 and for given number of zones p and subsequently choose the one that has the smallest value of the criterion.

4 Numerical experiments

The goal of numerical experiments is to compare proposed heuristic method with previously mentioned exact approach in terms of computational time and preciseness of the solution. To test behavior of proposed heuristic method we make a computational study on test data from selected region and compare the results obtained by this method with the solution of exact method. To be able to do it, we compare the effectiveness of heuristic and exact approach for a given value p.

Numerical experiments were performed on the data set created from the real public transportation network in the Zvolen Region in Slovak Republic. The stations in the networks are represented by the municipalities or part of municipalities. This network has 51 stations/municipalities respectively and is shown in the Figure 2. Black circles represent stations, the size of the circle represents approximate number of inhabitants and links represent existing connections of municipalities by public transportation.

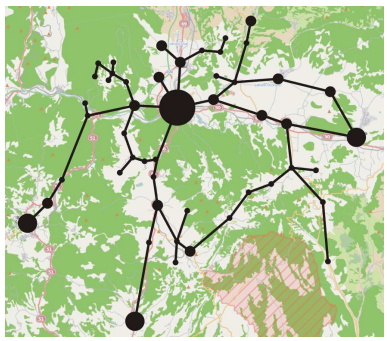


Figure 2: Test networks with 51 stations, Map source: *openstreetmap.org*

Current prices were calculated according to real prices depending on the distance for travelling by regional buses. The OD matrix was estimated using the gravity model as in [2], where the number of passengers between nodes i and j is calculated as follows:

$$\frac{b_i b_j}{d_{ij}}$$

where parameter b_i represents the number of inhabitants in the node i. The experiments were performed on a personal computer equipped with Intel Core 2 Duo E6850 with parameters 3 GHz and 3.5 GB RAM.

In the computational study we wanted to calculate optimal values of fare prices for selected values of parameter p. According to the current fare prices, we set the values of parameter f_1 from 0.3 to 0.9 with step by 0.1 and values of parameter f_2 from 0.1 to 0.6 with step by 0.1 for all the experiments. Table 1 represents results of nu-

merical experiments for the best values of average deviation calculated according to the (2) for given parameter *p*. Columns denoted as *ZP_ex* represent solutions obtained by exact method described in [5], columns denoted as *ZP_heur_dist* represent solutions obtained by proposed heuristic with the zone size calculation formula (10) and columns denoted as *ZP_heur_inhab* represent solutions obtained by proposed heuristic with the zone size calculation formula (11). Column *p_max* represent the given value of parameter *p*, columns denoted as *F** represent optimal values of objective function for best found values of parameters f_1 and f_2. Columns denotes as *Gap* represent values of gap between the best values of average deviation of particular heuristic method and optimal value achieved by exact method. In the Table 2 there is a comparison of computation times of all three approaches. Computation time includes the execution time of all 45 instances in each set.

p_max	*ZP_ex*	*ZP_heur_dist*		*ZP_heur_inhab*	
	*F**	*F**	**Gap [%]**	*F**	**Gap [%]**
4	9780.9	10651.2	8.89	10498.0	7.33
6	9655.7	10256.3	6.22	10786.1	11.71
8	9628.7	9989.0	3.74	10365.2	7.65
10	9230.2	9502.9	2.95	10199.0	10.49
13	8455.2	8905.2	5.32	9956.8	17.75
16	8266.1	8636.0	4.47	9795.3	18.49
20	7859.5	8233.2	4.75	9601.2	22.16
25	7991.0	8551.0	7.01	9068.5	13.48
30	8382.9	8760.2	4.51	8933.0	6.56

Table 1 Best values of average deviation

p_max	*ZP_ex*	*ZP_heur_dist*	*ZP_heur_inhab*
4	147.9	27.16	29.91
6	145.8	27.37	27.58
8	148.5	27.35	28.91
10	136.7	27.24	28.82
13	137.8	27.02	29.20
16	150.3	27.67	27.87
20	140.7	26.98	28.88
25	210.0	26.89	29.80
30	164.0	27.11	26.56

Table 2 Computation time in seconds

5 Conclusion

In the paper we described the mathematical model of the zone partitioning problem with given prices and number of zones with the average deviation criterion for counting zones tariff system. We proposed the heuristic method based on greedy approach with two different ways of calculating the zone size and performed the numerical experiments.

From the results of the numerical experiments in the Table 1 we can see, that the proposed heuristic method where we incorporate only distances to the zone size calculation *(ZP_heur_dist)* gives us more precise solutions than the approach with incorporation of number of inhabitants to the calculation of the zone size *(ZP_heur_inhab)*. Comparing computation time of exact method and both heuristic approaches we find that the both heuristic approaches are much faster than exact method.

In the future we would like to include also the OD matrix into the proposed heuristics, whereas the number of passengers has a significant impact on the value of the average deviation between the fair and the price for

traveling in such system. Numerical experiments will be also extended to the larger test network to study behavior of designed heuristics on larger problems.

Acknowledgements

This work was supported by the research grants VEGA 1/0339/13 "Advanced microscopic modelling and complex data sources for designing spatially large public service systems" and APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks". We would like to thank to "*Centre of excellence for systems and services of intelligent transport*" (ITMS 26220120050) for built-up infrastructure, which was used.

References

[1] Babel, L., Kellerer, H., 2001. *Design of tariff zones in public transportation systems: Theoretical and practical results.* Technical report, Faculty of Economics, University of Graz, Austria

[2] Cenek, P., Janacek, J., Janosikova, L., 2002. Location of Transportation Districts at Modeling of Tranportation Processes in a Region. *Communications - Scientific Letters of the University of Zilina* **4**, No. 1-2, 5-9.

[3] Hamacher, H. W., Schöbel, A.: Design of Zone Tariff Systems in Public Transportation. *Operations Research* **52** (2004), 897-908.

[4] Koháni, M.: Designing of zone tariff in integrated transport systems, In: *Communications: scientific letters of the University of Žilina* **15** (2013), 29-33.

[5] Koháni, M.: Zone partitioning problem with given prices and number of zones in counting zones tariff system, *In: SOR ´13: proceedings of the 12th International Symposium on Operational Research: Dolenjske Toplice, Slovenia* (2013), 75-80.

[6] Müller, S., Haase, K.., Klier, M., 2013. Revenue Maximizing Tariff Zone Planning in Public Transport, In: *Book of abstracts of XXVI EURO - INFORMS Joint International Conference*, Rome, 335.

[7] Palúch, S.: On a fair fare rating on a bus line. In: *Communications: scientific letters of the University of Žilina* **15** (2013), 25-28.

[8] Schöbel, A.: *Optimization in Public Transportation: Stop Location, Delay Management and Tariff Zone Design in a Public Transportation Network*, Springer, 2006.

[9] *XPRESS-Mosel User guide*. Fair Isaac Corporation, Birmingham, 2012.

Dynamic transportation problem with partial limitations of material for each participant of the transport

Pavel Kolman[1]

Abstract. The paper will deal with the issue of material distribution between suppliers and customers, the optimization criterion is to minimize the total traveling time. It is a material which is perishable and therefore the speed of transport has a priority. In the mathematical model, to each participant is given both: its daily transportation limits (they may vary during the week), and also total limits of the transported material. After the exhaustion of the shipped (received) material, the not supplier (customer) does not participate on the transport. The total amount of the offered and requested material may not be consistent, the transportation is completed after fulfillment of all the requirements, or by remaining of last transportation participant. Shipment time and delivery time do not correspond. The optimization criterion is to minimize total transportation time of the material.

Keywords: optimization, dynamic transportation problem, time minimization

JEL classification: C44, C61
AMS classification: 46N10, 90C05, 90C08

1 Introduction

Most transport problems, such as [2], solves transport to minimize transportation costs, what is understandable. In life, there are situations where the time criterion (ie minimizing the duration of transport) has a higher priority than the cost criterion. This are situations where the transported material is perishable (this type of problems do also indirectly minimizes costs), respectively where the rapid transport of the material can help to save human lives and minimize the considerable material damage, such as natural disasters.

There exists numerous of used models, but often is needed to design specific model for described situation, such as the calamity timber problem, described in [1, 5]. Described model solves the time minimization for a total amount of transported material, the transport of individual participants are limited only by the amount of material transported daily (total sent/received quantity for each participant of transportation is not set). The model was compliant for the collection of calamity wood, but there are situations where the participants of transport are limited by their daily, and the overall limits of transported material. With this modification it was necessary to design a fundamentally different model, including the method of its solution. Building the model and its solution is described in this paper.

2 Material and Methods

This section will describe the problem and its comparison with classical transportation problem, and will subsequently be described differences and the reasons why in this model can not be used classical transportation problem.

Participants of transportation problem are suppliers and customers. Let us therefore m suppliers $D_1, \ldots, D_m$ and n customers $O_1, \ldots, O_n$. Their job is to transport a homogeneous material from suppliers to customers. Transportation time between each supplier and the customer is known. These requirements are consistent with the classical transportation problem, as described for example in [2].

[1]Mendel University in Brno, Faculty of Business and Economics, Zemědělská 1, 603 00 Brno, xkolman@mendelu.cz

However, compared to the previous model, there are the following differences. The first is that there is an optimization criterion for the shortest transportation time. Partial transportation times between suppliers and customers are (without loss of generality) positive integers. Another difference is that in this type of model, are given both overall and daily capacities of suppliers respectively requirements of customers. Participants of the transport are therefore limited to a maximum daily capacity of materials that are able to send, or requirements that are able to accept. This may vary on individual days, and after some time repeats regularly. Time after which are the capacities of suppliers resp. requirements of customers regularly repeated (eg a week), will be called a time period. The time period (typically a week) consists of time units (days). Furthermore, it is determined the total quantity of material to be transported. During transport, the customer cannot accept more material than he can handle in the appropriate day.

In the context of the balanced problem is given the requirement that at least one side of the transport shall be fully satisfied. In case that the sum of the capacities of suppliers is less than the sum of the requirements of customers, all the material will be taken away, but some of the customers will not be satisfied. Otherwise, it will not be taken away all the available material, but all customers will be satisfied. From this point of view, the problem will always have a feasible (and thus the optimal) solution. To solve the described model it will be used the author's own optimization unit using simplex method, described in [3].

3 Results and Discussion

This section describes the building of a model and its solution. Because we are looking for an optimal solution in time, the problem can be called a dynamic transportation problem (DTP). However, because the abbreviation DTP is already longer used in the model without the overall limitations of individual suppliers and customers [5], will be used for the model abbreviation DTPN.

The idea of building and solution of a mathematical model is simple and will realized in two steps. The first step is (algorithmically) quick finding a feasible solution of the model, having the time of transport as far as the nearest to the optimal solution. In the other steps will be iterative shortening of total transportation time, until finding the optimal solution of the model.

3.1 Building and solution of the model designed to find a feasible solution

The main reason for building a model is to find a feasible solution, which will be subsequently improved. The idea of finding a feasible solution is relatively simple: each day will be sent the maximum amount of material, so that all constraints are satisfied. At the constrants of suppliers is not a problem, because the material is just send in the same day. For customers, the requirements will be reduced by the undelivered material sent from suppliers.

To maximize the amount of material sent daily we make a simple mathematical model of k-th transport day. It will contain a total of $m \cdot n$ variables, where m is the number of suppliers and n of customers. The $x_{i,j,k}$ variables will describe the amount of material sent by i-th supplier j-th customer in k-th day of transport. All coefficients of the objective function $c_{i,j}$ will have a value of 1 and the problem will be maximization type (we are looking for a solution that maximizes the amount of material sent to the selected date), ie

$$z_{\max} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{i,j,k}. \tag{1}$$

The model will occur following types of constraints:

- From each of the suppliers should not be sent more than its maximum daily limit, a total of m restrictive conditions.
- None of the customers in one day can be delivered more material than its maximum limit for the day. Since the various transport time (it can take up to r time units), we need no more than $n \cdot r$ restrictive conditions, considering only the constraints that may actually occur.

- The overall customers limits cannot be exceeded, total of n restrictive conditions. (Restrictive conditions describing the exceeding of overall limits of suppliers are not taken into account.)

The first type of constraints is m restrictive conditions relating to the suppliers. Each of them expresses that it can not be sent more material than the limit of k-th day, therefore, limit of i-th supplier will be written in the form

$$x_{i,1,k} + x_{i,2,k} + \cdots + x_{i,j,k} + \cdots + x_{i,n,k} \leq b_{i,\star,k}, \tag{2}$$

where $b_{i,\star,k}$ is daily limit of i-th supplier in the k-th day of transport and the symbol "$\star$" means any customer. Here it should be noted that if the total outstanding amount of material is less than the applicable daily limit of i-th supplier $B_{i,\star,k}$, the right-hand side of this constraint will be replaced by that value.

Constraint on the demand side are little bit more complicated, because their daily limits may already be reduced by the material that was sent to the relevant customers in previous days from one of the suppliers (with the exception of the transport start, when this situation does not occur). Therefore, it is firstly necessary to determine the maximum transporting time r, according to the formula

$$r = \max\{\mathrm{d}(i,j)\}, \tag{3}$$

where $\mathrm{d}(i,j)$ is the duration of transport between i-th supplier ($i = 1,\ldots,m$) and j-th customer ($j = 1,\ldots,n$).

The material was sent in one day, but to the customers will be delivered within a maximum r days. Therefore, we will compile a maximum of $n \cdot r$ constraints describing these delivering in these days. In fact, they will be build only the limitations in which the material sent in the k-th day can be actually delivered.

Constraint describing the limit of j-th customer with the delivery of material after $k+p$ days of the dispatch, where $1 \leq p \leq r$, will be expressed as:

$$a_{1,j} \cdot x_{1,j,k} + \cdots + a_{i,j} \cdot x_{i,j,k} + \cdots + a_{m,j} \cdot x_{m,j,k} \leq b_{\star,j,k+p} - \sum_{l=1}^{r} x_{i,j,k+p-l}, \tag{4}$$

where $x_{i,j,k+p-l}$ is a material supplied from the i-th supplier to the j-th customer l days before the date of delivery $k+p$ and structural coefficients $a_{i,j}$ take values

$$a_{i,j} = \begin{cases} 1 \text{ if } \mathrm{d}(i,j) = p, \\ 0 \text{ otherwise.} \end{cases} \tag{5}$$

It makes sense to prepare constraints for the values of p that occur in the matrix of the duration of transport.

The last type of constraints are the total limits of the customers. Restrictive conditions limiting the total amount of material sent from j-th customer in k-th day will be prepared according to the formula:

$$x_{1,j,k} + \cdots + x_{i,j,k} + \cdots + x_{m,j,k} \leq B_{\star,j} - \sum_{q=1}^{k-1} x_{\star,j,q}, \tag{6}$$

where the expression $B_{\star,j}$ represents the total requirements of j-th customer and expression $\sum_{q=1}^{k-1} x_{\star,j,q}$ quantities of material sent from any supplier to j-th customer.

Solving the model for one day (one time unit) will affect the right-hand sides of the model in the next days (time units). The model is established and solved as long as all the available material is taken from suppliers, respectively. until all the material they customers want is delivered to them. Using this algorithm we obtain feasible solution of described problem.

3.2 Building a model solution designed to find the optimal solution

For the algorithm described here, the mathematical model for the entire transport will be build. The problem is that the time needed for transportation is not known. Therefore, we firstly find a feasible solution based on time needed for transportation and pursuant the obtained solution the mathematical model will be formulated.

The mathematical model will include a total of $m \cdot n \cdot (q+r)$ structural variables, where q is the number of days when the material will be dispatched, and r the maximum transport time between suppliers and customers. It is necessary that q be so great to find feasible (respectively optimal) solution of entire transport. This can be ensured eg using the algorithm described in subsection 3.1.

Binding constraints of the model will again be of several types, there will be a total of $m \cdot q + n \cdot (q + r) + m + n + 1$ constraints.

The first $m \cdot q$ of constraints quarrantees the maximum daily amount of sent material. Constraint describing the i-th supplier in the k-th day, when the material shipped, we write according to the formula

$$\sum_{j=1}^{n} x_{i,j,k} \leq b_{i,\star,k}. \tag{7}$$

Another $n \cdot (q+r)$ constraints, guarantees not exceeding the maximum daily amount of delivered material for j-th customer in k-th day will be written as

$$\sum_{i=1}^{m} x_{i,j,K} \leq b_{\star,j,k}, \tag{8}$$

where K is the time of the material dispatch, calculated according to the formula

$$K = k - \mathrm{d}(i,j). \tag{9}$$

Followed building m constraints ensuring that within the entire period it will not be sent more material than the supplier has, ie, not exceeding the total limit of i-th supplier we writte down as

$$\sum_{j=1}^{n} \sum_{k=1}^{q+r} x_{i,j,k} \leq B_{i,\star}, \tag{10}$$

where $B_{i,\star}$ is the total capacity of the i-th supplier.

Another n constraints guarantees that during the entire period the requirements of the supplied material will not exceed. For j-th customer will be constraint written in the form

$$\sum_{i=1}^{m} \sum_{k=1}^{q+r} x_{i,j,k} \leq B_{\star,j}. \tag{11}$$

The last constraint ensures transporting a given amount of material, according to the formula

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{q+r} x_{i,j,k} = \min \left\{ \sum_{i=1}^{m} B_{i,\star}, \sum_{j=1}^{n} B_{\star,j} \right\}. \tag{12}$$

With the exception of the objective function, the model is build. The aim is to minimize the time of dispatch of the material. The idea of finding the optimal solution and optimality test is as follows: The model consists of $q+r$ days, but the material is sent within the first q days. Objective function has to be maximized. Variables representing the first $q-1$ days will have the objective function coefficients 1 and variable representing the q-th to $q+r$-th day coefficients 0. This problem we solve. If it has a feasible solution, while the value of the objective function is equal to the expression

$$z^*_{\max} = \min \left\{ \sum_{i=1}^{m} B_{i,\star}, \sum_{j=1}^{n} B_{\star,j} \right\}, \tag{13}$$

then the number of q days shipment **is not optimum** because there exists feasible solution that sends all the material in less than q days (and that was just found). In case the solution is feasible, while the inequality

$$z^*_{\max} < \min\left\{\sum_{i=1}^{m} B_{i,\star}, \sum_{j=1}^{n} B_{\star,j}\right\}, \tag{14}$$

then the number of q days of shipment **is optimum**, because there is no feasible solution, which would be able to transport the material in less than q days. This is because the coefficients in the objective function have a value of 1 for the variables representing the first $q-1$ days and sum of the transported amount of material is less than it should be transported.

If the found solution is not optimal, ie is valid expression (13), we perform the following adjustment of the model: coefficients representing the q-th to $q+r$-th day we set as prohibitive rate $-M$, coefficients of $q-1$-st day we set as 0. For the modified model, the obtained solution becomes infeasible, so we re-solve it. We continue like this until we find a solution that satisfies the expression (14). This solution is optimal, the time required to send all the material is equal to the index of sending variables having the coefficient of the objective function zero.

If we want to minimize the delivery time instead of the time of dispatch, we will proceed by analogy, the only changes in the numbering of the coefficients of the objective function will be performed with regard to the date of delivery of the material, instead of sending.

4 Conclusion

The described method was designed and used in solving the dissertation thesis [4]. The paper describes the method of solution, but only in terms of model building and finding the optimal solutions. The method is not dealt with regard to the used algorithms. This part is one of the most important, since the choice of algorithms and their use greatly affects the relationship between the size of the problem and the time required to find an optimal solution (ie the time complexity of used algorithms). With effective programming, the time complexity of the proposed method is sufficient to obtain optimal solutions in reasonable time. However, the detailed complexity analysis is beyond the scope of this article.

References

[1] Bulantová, P.: *Využití lineárního programování při optimalizaci dopravy.* Bachelor thesis. Brno, MZLU v Brně, 2008.

[2] Jablonský, J. *Operační výzkum.* Praha, VŠE, 2001. ISBN 80-245-0162-7.

[3] Kolman, P.: Implementation of blending problem algorithm into information system. *In Proceedings of the 30th International Conference Mathematical Methods in Economics 2012.* Karviná, Silesian University in Opava, School of Business Administration in Karviná, 2012, 438–442. ISBN 978-80-7248-779-0.

[4] Kolman, P.: *Modely dopravních úloh minimalizující dobu přepravy.* Dissertation thesis. Brno, Mendelova univerzita v Brně, 2012.

[5] Kolman, P., Střelec, L.: Time minimizing transportation of calamity fallen timber. *In SIMOS, T E. 11th International Conference of Numerical Analysis and Applied Mathematics 2013: ICNAAM 2013, AIP Conference Proceedings.* Melville, New York, American Institute of Physics, 2013, 1843–1846. ISBN 978-0-7354-1184-5.

The impact of China's economic rise on the Czech Republic's motor vehicle exports

Veronika Končiková[1]

Abstract. The surge of China as a global exporter raised numerous questions about how it might influence other economies. At first, literature focused on developing economies, but the shift of China's export structure towards more sophisticated products might bring a competition effect even for the developed countries.

In this paper we examine whether Czech Republic's most important export commodity – the motor vehicles - is influenced by China's growing importance in the global trade. Using the gravity models we examine whether Czech car industry suffers from Chinese competition in its export markets. The paper is organized as follows: the first chapter briefly introduces the main development in the China's export structure and provides a brief literature review of this topic, the second chapter describes the methodology and model used in our analysis and the following chapter presents our main findings. The conclusion and possibilities for further research in this area are offered at the end.

Keywords: Trade, economic performance, automotive industry, exports, China, Czech Republic.

JEL Classification: C33, F10
AMS Classification: 90C15

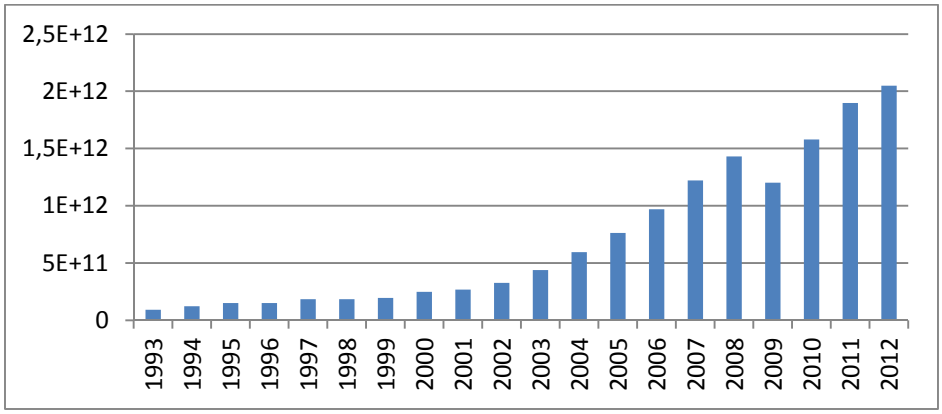
1 China's exports

In the last decades China underwent important economic changes which lead to remarkable changes in Chinese economy. One of the biggest changes China experienced was its opening to the World. As noted in [12], while China's exports were exceptionally small in the 60's, they accounted for approximately 30% of its GDP at the beginning of the new century. In this section we first present the development of China's export and in the second part we focus on the literature emphasizing China's impact on other exporters.

1.1 The rise of China's exports and its changes over time

China's entrance as global trade power in the second half of the 20th century was almost unprecedented and China's exports continued to growth in the last years. As Figure 1 demonstrates only between the years 1993 and 2012 Chinese exports grew almost 17 times bigger and according to [20] nowadays Chinese share in world total exports is 11.13%.

Figure 1 Total Chinese exports (in USD)



Source: UN Contrade

[1] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno-Pisárky, Czech Republic, e-mail: 207077@mail.muni.cz

It is important to notice that Chinese striking growth in exports was in the recent years accompanied by an important change in export structure. As mentioned in [18] the proportion of high-tech exports grows over time on the expense of the low-tech exports. However, in [21] it is argued that this phenomenon is mainly caused by the processing trade when the high-tech imports are assembled in China. These assembling operations are of very low added value. In contrary to these findings in [1] we find support for the delinking of Chinese exports and imports. This funds the idea that China's production is more sophisticated than what its income level would suggest.

The structure of Chinese exports is crucial to our analysis. When China was exporting only low-tech products, it was very unlikely that it would impose a threat to the exports of developed economies. In [17] Vlčková argues that China is still not able to intimidate the position of western European countries as technological leaders. However in [13] it is showed that the exports of China and OECD countries overlap and this trend is increasing over time. This overlap might indicate that even developed countries might be fighting for export markets with China.

Furthermore Chinese export of motor vehicles has been growing importantly in the last years. Nataraj and Tandon in [10] have identified motor vehicles as being one of the most dynamic Chinese exports to the world during the period 2001-2006. Additionally in [9] it is proved that motor vehicles were the most rapidly growing Chinese export commodity to the Czech market in the recent years.

1.2 The impact of China's export growth on other economies

When seeing the changing volume and structure of Chinese exports, one must ask what impact this might have on other economies. Unsurprisingly, first such research was done concerning the impact of Chinese exports on developing countries. At an earlier stage of its development China was primarily exporting low-value-added goods which are mostly produced in developing countries.

One of the first empirical researches in this topic was conducted by Eichengreen et all. in [4]. Authors were assessing the impact of China on Asian economies. They found that there is a different effect on exporters of consumer goods and on the exporters of consumption goods, hence there is different influence on more and less developed Asian countries. While the more developed countries in Asia saw their exports grow thanks to Chinese increasing demand, the less developed countries exports shrunk due to the Chinese competition in its export destinations. In [6] Fu et all. further develop their idea and try to examine what impact China has on low and middle income countries' export prices.

The recent shift in Chinese export structure has urged the literature focusing on the developed countries. In [7] Giovanetti et all. focuses on four European countries (France, Germany, Italy and Spain) and they found that even in the case of developed countries the export competition is different across the countries. Exporters of traditional goods such as Italy face much higher competition from China than exporters of more advanced goods such as Germany. Finally, Flückiger and Ludwig in [5] examine what impact has Chinese growing competition on the OECD countries. They conclude that Chinese emergence influences the advanced economies on various ways what supports our hypothesis that Czech exports in vehicles might be affected by China's rise.

In our paper we will follow these previous researches and try to assess the impact of Chinese exports in one particular commodity – Motor vehicles for the transport of persons (7812) – on the Czech Republic's exports. As in all of the above mentioned studies, gravity model will be used to evaluate this impact.

2 Gravity models

Already for decades gravity models have been widely used to analyze the external flows (such as trade) among countries. Gravity models were first proposed in the 60's by Tienbergen in [15] and Pöyhönen in [11], but not until the 80's were they given a proper theoretical framework proposed in [2].

The principal idea of gravity models is taken from the natural sciences: as the name suggest, this econometric approach follows the idea of Newton's law of universal gravitation. While Newton's law suggests that the force between masses is positively determined by the masses of objects and negatively by the distance between them, the gravity models used in economic research propose that this law can be transmitted into analysis of international flows such as trade, investment or migration.

According to gravity models of international trade, the flows between countries are determined by the "mass" of a country which is usually believed to be well represented by the GDP. Another determinant of trade is the distance between countries where we expect negative correlation. The intuitiveness behind this simple model is very clear: according to trade theories we expect countries to trade more when they have larger markets and we anticipate that they will trade less when transaction and transportation costs take place.

Other explanatory variables might be added to the model according to what analysis is the model used for. Gravity models have frequently studied the impact of free trade zones, currency unions or other trade-related economic policies on these countries' trade. In the Czech context gravity models have been used for example in [14] to examine the determinants of Czech agricultural exports or in [8] to analyze the impact of state subsidies and export credit on Czech exports. In this paper, we will use the gravity model to examine the impact of the Chinese exports on the Czech exports.

2.1 Methodology

As stated previously our model is based on the gravity model and the following equation is used:

$$EX_{it} = \alpha + \beta_1 CH_{it} + \beta_2 GDP_{it} + \beta_3 GDP_{ct} + \beta_4 D_i + \beta_5 B_i \tag{1}$$

Where EX_{it} stands for Czech exports to country *i* in time *t*, CH_{it} stands for Chinese exports to country *i* in time *t*, GDP_{it} stands for gross domestic product of importer country in time *t*, GDP_{ct} stands for gross domestic product of Czech Republic in time *t*, D stands for distance between Prague and the capital city of the country *i* and B stands for a dummy variable for common borders with Czech Republic. All the parameters are the same for all the countries involved in our model.

We use the panel model in our analysis in gretl because our model includes time series as well as cross-sectional data. Because our model includes data which are constant over time (such as the distance between the capital cities) the fixed effects cannot be used because these variables would be omitted. That is why we opted for the use of random effect models in our paper.

2.2 Data

The data used in our dataset come from three different sources: [3], [16] and [19]. The data for trade have been collected from UN Comtrade database. The Standard International Trade Classification was used in the third revision. Despite the availability of the revision four, we elected the use of revision three due to longer time series available. Our dataset obtains observation of exports from Czech Republic and China to 195 countries which together with the time period going back to the formation of the Czech Republic in 1993 allows us for up to 3380 observations in our models. The data are available in current US dollars. As mentioned previously our study will focus only on the road vehicles and therefore only export data for one commodity were used - Motor vehicles for the transport of persons (7812). In our model first differences of exports have been used to avoid problem with non-stationarity.

The information about gross domestic product was taken from the Worldbank's databank. The data represent the total gross domestic product of the selected countries in current US dollars. We opted for these variables in order to stay consistent with the data collected for the exports. The first differences of the Czech and Chinese GDP were used.

The data about distance between the capital cities were downloaded from the CEPII – a French research center in international economics which provides numerous gravity-related datasets. As mentioned previously the most commonly used indicator for distance between countries was used – the distance between their capital cities.

3 China's rise and Czech car exports

The impressive growth of overall Chinese exports and its shifting export structure suggest that China's growing influence might have an important impact on countries' economies and even more importantly on the small open economies such as Czech Republic. The findings in [9] and [10] advocate that Czech Republic's export in one of its main export commodities – personal motor vehicles – can face a competition from China in its export markets. Hence we formulate a hypothesis that Czech car export might be decreasing in the countries where fierce competition from China emerges. We will test this hypothesis using the equation (1) and therefore expect a negative sign for the coefficient for Chinese exports to country *i*.

As suggested previously we first tested the model using the random effects. The Table 1 summarizes the main findings:

variable	coefficient	p-value	
	4.74691e+06	0.00828	***
CH_{it}	0.289277	<0.00001	***
GDP_{it}	3.85855e-05	0.00040	***
GDP_{ct}	9.1675e-05	0.01950	**
D	-647.782	0.00747	***
B	5.8214e+07	<0.00001	***

Table 1 Gravity model using random effects

However the Hausman test indicates that the use of fixed effects would be more suitable for our model. Therefore we repeat the analysis using the fixed model effects despite the fact that we will lose some of the important regressors. Table 2 summarizes our results:

variable	coefficient	p-value	
	2.24326e+06	0.00150	***
CH_{it}	0.280167	<0.00001	***
GDP_{it}	3.45938e-05	0.00532	***
GDP_{ct}	9.52787e-05	0.01548	**

Table 2 Gravity model using fixed effects

In the model using fixed effects we find out that the results did not change significantly from the previous estimation using the random effects. Therefore despite the fact that our model is not perfect, we can derive some important suppositions.

Both of the models offer similar results therefore our interpretation is related to both of them. We observe that the traditional gravity variables have the expected signs of their coefficients as suggested by economic theory. The gross domestic product of exporting as well as importing countries positively affects the exports of Czech vehicles. However, note that this impact is relatively small. This suggests that other variables must play a more important role. When taking into account the random effect model, we can see that distance and the border dummy play an important role in determining the Czech vehicle exports. This finding is also in accordance with the theoretical predictions. The distance and border dummy should represent a proxy for the transaction (most importantly the transportation cost). The bigger the distance, the smaller are the exports. It is also logical to assume that the bordering countries will trade more not only thanks to the geographical proximity, but also due to the historical ties and shared cultural predispositions.

Conversely, the variable of Chinese exports is different than suggested by our hypothesis. Moreover, it is indicated that there is a positive impact of Chinese exports on Czech exports in the field of personal motor vehicles. We can therefore conclude that China's rise as global trader and its growing exports in motor vehicles are, so far, not a threat to Czech Republic's exports of cars. This can be explained by Czech and Chinese cars not being perfect supplements as our hypothesis could suggest.

4 Conclusion and Discussion

In this paper we examined the effect of Chinese exports of motor vehicles on the Czech exports in the same commodity. Using gravity models and panel data analysis we found out that the Czech exports are positively influenced by its own GDP as well as by the GDP of the importing country. In addition, countries sharing the borders with Czech Republic import more Czech cars than others. The distance seems to be an important factor determining the export destination of Czech cars. As the logic suggests, the further is the destination, the less likely is Czech Republic export to this country. However we did not find any evidence for our hypothesis that Czech Republic could face competition from China in its export markets. It even seems that the rising Chinese exports to these countries enhance the Czech exports.

There might be several explanations to why in our model we do not observe a competition in this field between Czech Republic and China. The first and most likely is that the motor vehicles from China are not substitutes: they are different cars, with different quality, bought by different customers. The second explanation could be the changing nature of global trade. With division of trade into trade in vertical and horizontal specialization it

is very hard to track which of the countries was an actual producer of the good being exported using the available statistics.

This paper is a preliminary analysis for further research, which should not only include more export commodities, but should also embrace more European countries. The model used in this paper allows us to obtain a first glimpse on the basic relation between China's emerging car industry and Czech performance in this sector. Hence, the future research should conduct more detailed analysis using also Hausman-Taylor estimator and alike.

Acknowledgements

This research was supported by the grant Economic Policy in EU and Czech Republic No. MUNI/A/0811/2013, the grant was provided by Masaryk University.

References

[1] Amiti, M. and Freund, C.: *An Anatomy of China's Export Growth.* In FEENSTRA, W. – WEI, S. *China's Growing Role in the World Trade*. The University of Chicago Press, Chicago, 2010.
[2] Bergstrand, J. H.: The Gravity Equation in International Trade: Some Microeconomics Foundations and Empirical Evidence. *The Review of Economics and Statistics* **vol.67**, 3 (1985), 474-481.
[3] CEPII. 2014, UN. Accessed 10.4.2014 on the web page: <http://www.cepii.fr/CEPII/fr/cepii/cepii.asp>
[4] Eichengreen, B., Rhee, Y., Tong, H.: The Impact on the Exports of Other Asian Countries. *Review of World Economics* ***v*ol.143**, 2 (2007), 201-226.
[5] Flückiger, M., Ludwig, M.: Made in China: Export competition and structural changes in the OECD countries. *Working Paper* (2012)
[6] Fu, X., Kaplinsky, R., Zhang, J. The Impact of China on Low and Middle Income Countries' Export Prices in Industrial-Country Markets. *World Development* **vol. 40**, 8 (2012), 1483-1496.
[7] Giovanetti, G., Sanfilippo, M., Velucchi, M. The "China effect" on EU Exports to OECD markets – A focus on Italy. *Working Paper* (2011)
[8] Janda, K., Michalíková, E. and Potácelová, V.: Gravitační a fiskální modely státní podpory exportních úvěrů v České republice. *Politická Ekonomie* **vol. x**, 3 (2010), 305-325.
[9] Končiková, V.: Sino-Czech bilateral trade: structural resemblances and deviations. In *Zborník príspevkov zo 14. medzinárodnej konferencie: Medzinárodné vzťahy 2013: Aktuálne otázky svetovej ekonomiky a politiky*. (Lipková, Ľ. And Kucharčík, R., eds.). Ekonóm, Bratislava, 2013, 352-365.
[10] Nataraj, G. and Tandon, A.: China's changing Export Structure: A factor based analysis. *Economic and Political Weekly* **vol.66**, 13 (2011), 130-136.
[11] Pöyhönen, P. A.: Tentative Model for the Volume of Trade between Countries. *Weltwirtschoftliches Archiv* **vol.90**, 1 (1963), 93-99.
[12] Rodrik, D.: What's so special about China's exports? *China & World Economy* **vol.14**, 5 (2006), 1-19.
[13] Scott, P. K.: The relative sophistication of Chinese exports. *Economic Policy* vol.23, 53 (2008), 5-49.
[14] Ševela, M.: Gravity-type model for Czech agricultural export. *Agricultural Economics* **vol.48**, 10 (2002), 463-466.
[15] Tienbergen, J. *Shaping the World Economy: Suggestions for an International Economic Policy.* The Twentieth Century Fund, New York, 1962.
[16] UN Comtrade(United Nations Commodity Trade Statistics). 2014, UN. Accessed 10.4.2014 on the web page: <http://comtrade.un.org/>
[17] Vlčková, J.: Is China catching up? Technological sophistication of Chinese exports to European Union. *Acta Oeconomica Pragensia* **vol.20**, 3 (2012), 36-54.
[18] Wei, H. and Wang, X.: The trade structure of Chinese manufactured exports: 1999-2009. *American Journal of Economics and Business Administration* vol.4, 4 (2012), 197-206.
[19] Worldbank. 2014, UN. Accessed 10.4.2014 on the web page: <http://databank.worldbank.org/data/home.aspx>
[20] WTO. 2014, UN. Accessed 10.4.2014 on the web page: <http://stat.wto.org/CountryProfile/WSDBCountryPFView.aspx?Country=CN&>
[21] Xing, Y.: The People's Republic of China's High-Tech Exports: Myth and Reality. *ADBI Working Paper Series* (2012).

Minimal Risk Portfolios under SSD efficiency constraints

Miloš Kopa[1]

Abstract. This paper deals with new types of optimization problems when minimizing a risk of a portfolio under a condition on portfolio mean return and over portfolios which are classified as efficient with respect to second-order stochastic dominance (SSD) criterion. These problems can be seen as generalizations of classical mean-risk models where a risk measure is minimized under condition on portfolio mean return. The crucial condition on the second order stochastic dominance efficiency is expressed in terms of existence of "optimal" utility function which obeys SSD rules. It means that new problems find portfolios having minimal particular risk measure (variance, Value at Risk, conditional Value at Risk), with at least minimal required mean return and being the optimal solution of maximization expected utility problems for at least one non-decreasing and concave utility function. This study reformulates these new problems in linear, nonlinear, mixed-integer programs. Moreover, using US stock market data, paper compares the classical mean-risk frontiers with optimal portfolios of the new problems with SSD efficiency constraints.

Keywords: stochastic dominance, portfolio efficiency, risk measuring.

JEL Classification: D81, G11
AMS Classification: 91B16, 91B30

1 Introduction

Stochastic dominance (SD), first introduced in Quirk and Saposnik [23], Hadar and Russell [7] and Hanoch and Levy [8], is a useful approach to decision making under risk when only partial information about the decision maker's risk preferences (utility function) is available. SD imposes general preference restrictions without assuming a functional form for the decision maker's utility function. It considers a class of utility functions consistent with the decision maker's preferences. The concept is used in numerous empirical studies and practical applications, ranging from agriculture and health care to financial management and public policy making; see, for example, the extensive survey in the text book of Levy [14].

The SD rules of order one to four are particularly interesting, because they impose (in a cumulative way) the standard assumptions of non-satiation, risk-aversion, prudence and temperance, which are necessary conditions for standard risk aversion. Moreover, the most popular SD criterion is the second-order SD (SSD). It assumes only non-satiation and risk aversion for the preferences. There exist several alternative formulations of this criterion in terms of concave utility functions, cumulated distribution functions, second quantile functions, lower partial moments, Lorenz curve, cumulative returns and Conditional Value at Risk see, for example, Levy [14], Ogryczak and Ruszczyński [18], Kopa and Chovanec [10]. The traditional approach to SSD analysis uses crossing algorithms that check if the cumulated empirical cumulative distribution functions of a pair of choice alternatives cross at some point. Unfortunately, this approach becomes computationally non-tractable when a given choice alternative is compared to all possible portfolios created from the given choice alternatives. This problem arises typically in the context of portfolio selection problems where the investor can fully diversify across financial assets and hence infinitely many choice alternatives exist.

Therefore Post [20] and Kuosmanen [13] developed linear programming tests for testing if a given portfolio is SSD efficient relative to all possible portfolios formed from a set of assets. Using the formulation in terms of concave utility functions and the first-order condition for portfolio optimization, Post [20] derives a computationally efficient LP test. Using the formulation in terms of cumulative returns and the majorization principle, Kuosmanen [13] derives a test that identifies another, SSD efficient portfolio that dominates the evaluated portfolio (if the latter is inefficient). This test involves solving two linear problems; one for a necessary condition and one for a complementary sufficient condition. As it was shown in Ogryczak and Ruszczyński [18], SSD criterion can be expressed in terms of second quantile functions or CVaRs. These expressions were applied to portfolio efficiency testing in Kopa and Chovanec [10]. Recently Lizyayev [15], and Post and Kopa [21] derived

[1] Charles University in Prague, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague, Czech Republic,
e-mail: kopa@karlin.mff.cuni.cz

other formulations as modifications of the previous tests that are useful in some situations. Finally, Kopa and Post [12] introduced a general SSD efficiency test. All the previous tests can be seen as particular (special) examples of the general test.

This paper introduces and analyzes a new type of optimization problem which combines risk minimizing and mean return maximizing objectives with portfolio efficiency constraints. Risk minimizing problems under condition on mean return are well known in the decision making theory (or portfolio selection theory) as "mean-risk" models. The first portfolio selection problem in the mean-risk formulation was introduced by Markowitz [17]. The model jointly focuses on maximizing expected return and minimizing variance of the portfolio, where variance serves as a measure of risk. In last 60 years, several other risk measures were developed in the theory of mean-risk models, for example, Value at Risk (VaR) or Conditional Value at Risk (CVaR), see Pflug [19], Rockafellar and Uryasev [24]. Later on, the mean-risk models were enriched by stochastic dominance relations. Dentcheva and Ruszczyński [3] introduced portfolio selection models that minimize an objective on the set of portfolios that dominates a benchmark with respect to the second order stochastic dominance.

Our risk minimizing models under SSD efficiency constraints can be seen as another modification of a classical mean-risk model. Instead of constraints in the form of SSD relation (Dentcheva and Ruszczyński [3]) we rather consider as feasible points only portfolios which are SSD efficient. This new approach guarantees that the optimal portfolio is efficient with respect to second-order stochastic dominance relation. This is an important feature of these problems, because classical mean-risk problems, in general, need not to identify a SSD efficient portfolio, as it was criticised many times, see Levy [14]. It means that it can easily happen that the optimal portfolio of a mean-risk problem is an optimal choice for no non-satiated and risk averse investor.

The rest of the paper is structured as follows. Section 2 introduces the basics of stochastic dominance and risk measures. It is followed by formulation of SSD efficiency constraints in Section 3. This section also derives the new-type problems for the particular risk measures as linear, non-linear or integer programs. Section 4 applies the theoretical formulations to US stock market data and compares new efficient frontier with classical mean-risk frontier. Finally, Section 5 concludes the paper and presents some remarks for the further possible modifications or improvements.

2. Preliminaries

We consider a random vector $\mathbf{r} = (r_1, r_2, \dots, r_n)'$ of returns of n base assets. We assume that the returns have discrete probability distribution. The distribution is described by T equiprobable scenarios. The returns of the assets for the various scenarios are given by

$$X = \begin{pmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^T \end{pmatrix}$$

where $\mathbf{x}^t = (x_1^t, x_2^t, \dots, x_n^t)$ is the t-th row of matrix X. Portfolio weights $\mathbf{w} = (w_1, w_2, \dots, w_n)'$ are assumed to be from set: $W = \{\mathbf{w} \in R_+^n: \mathbf{1}'\mathbf{w} = 1\}$, that is, no short positions are allowed and only a budget constraint is imposed. This choice of portfolio feasibility set is very common, however, one can easily replace it by any bounded polyhedral set and do all derivations and computations with this more general choice.

For any portfolio $\mathbf{w} \in W$, let $(X\mathbf{w})^{[k]}$ be the k-th smallest element of $(X\mathbf{w})$, i.e.

$$(X\mathbf{w})^{[1]} \leq (X\mathbf{w})^{[2]} \leq \cdots \leq (X\mathbf{w})^{[T]}.$$

Definition 1: *We say that a portfolio $\boldsymbol{v} \in W$ dominates another portfolio $\boldsymbol{w} \in W$ with respect to the **second** order stochastic dominance if and only if $Eu(\boldsymbol{rv}) \geq Eu(\boldsymbol{rw})$ for all **concave** utility functions u provided the expected values are finite.*

The second order stochastic dominance relation can be verified by comparing either twice cumulative distribution functions or second quantile functions or conditional values at risk, see for example Kopa and Chovanec [10]. Alternatively, following Kuosmanen [13] and Luedtke [16], one can express the SSD relation using a double stochastic matrix Q.

Our risk minimizing models use three different measures of risks for returns of portfolio $\boldsymbol{w} \in W$: variance, Value at Risk (VaR) and conditional Value at Risk (CVaR) which can be computed as follows:

$$\sigma^2(\mathbf{rw}) = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}^t\mathbf{w} - \frac{1}{T}\sum_{s=1}^{T}\mathbf{x}^t\mathbf{w}\right)^2,$$

$$\mathrm{VaR}_\alpha(-\mathbf{rw}) = (-X\mathbf{w})^{[s]} \text{ where } \frac{s-1}{\mathrm{T}} < \alpha \leq \frac{s}{\mathrm{T}},$$

$$\mathrm{CVaR}_\alpha(-\mathbf{rw}) = \min_{\mathrm{y},\mathbf{z}} \mathrm{y} + \frac{1}{(1-\alpha)T}\sum_{t=1}^{T} z_t \text{ s.t. } z_t \geq -\mathbf{x}^t\mathbf{w} - \mathrm{y}, z_t \geq 0, t = 1, \dots, T.$$

While variance and conditional Value at Risk can be directly combined with mean return constraints or other constraints, Value at Risk needs a reformulation using auxiliary binary variables $\delta_\mathrm{t}, \mathrm{t} = 1, \dots, \mathrm{T}$:

$$\mathrm{VaR}_\alpha(-\mathrm{rw}) = \min_{\delta,\gamma} \gamma$$

$$\mathrm{s.t.}\ \gamma + \mathrm{M}.\delta_\mathrm{t} \geq -\mathrm{x}^\mathrm{t}\mathrm{w}, \mathrm{t} = 1, \dots, \mathrm{T},$$

$$\sum_{\mathrm{t}=1}^{\mathrm{T}} \delta_\mathrm{t} = \lfloor(1-\alpha)\mathrm{T}\rfloor,$$

$$\delta_\mathrm{t} \in \{0,1\}, \mathrm{t} = 1, \dots, \mathrm{T},$$

where $\lfloor(1-\alpha)\mathrm{T}\rfloor$ denotes the highest integer number not exceeding $(1-\alpha)\mathrm{T}$ and M is the sufficiently large constant. Risk minimizing problem with VaR is more computationally demanding than the problem with variance or CVaR because of the auxiliary binary variables. While mean-risk problem with CVaR is a linear program, employing variance one gets the quadratic program and, moreover, if VaR is considered the formulation of the model leads to (even more complicated) mixed integer linear program.

3. SSD efficiency conditions

Since decision maker can combine assets into portfolios, an efficiency approach based on SSD relation is more of interest than just pairwise comparisons of two given portfolios (see Definition 2). We follow Post and Kopa [21] in the definition of SSD efficiency and in derivation of a necessary and sufficient conditions for SSD efficiency of a given portfolio. For simplicity, let $\mathrm{y}^t = \mathrm{x}^\mathrm{t}\mathrm{w}, t = 1,2, \dots, T$ and assume that scenarios are ordered by their ranking with respect to the evaluated portfolio: $y^1 \leq \cdots \leq y^\mathrm{T}$.

Definition 2: *An evaluated portfolio* $\mathbf{w} \in W$ *is efficient in terms of the second order stochastic dominance, relative to all feasible portfolios if it is the optimum for some concave utility function, that is:*

$$\sum_{t=1}^{T} u(y^t) \geq \sum_{t=1}^{T} u(\mathrm{x}^\mathrm{t}\boldsymbol{v}) \quad \forall \boldsymbol{v} \in W \Leftrightarrow$$

$$\sum_{t=1}^{T} u'(y^t)(y^t - x_j^t) \geq 0 \quad j = 1, \cdots, n.$$

The definition follows from the Karush-Kuhn-Tucker first-order condition for selecting the optimal combination of prospects. This formulation was first introduced by Post [20] for SSD ($N = 2$) and applies also for higher-order criteria ($N > 2$), see Post and Kopa [21] for more details.

Alternatively a given portfolio w ∈ W is inefficient with respect to the second order stochastic dominance (SSD inefficient) if there exists portfolio v ∈ Wwhich dominates w with respect to the second order stochastic dominance relation. If such portfolio does not exist then w is SSD efficient.

Theorem 1 (Post and Kopa [21]): *An evaluated portfolio* $w \in W$ *is efficient in terms of the second order stochastic dominance, relative to all feasible portfolios if and only if there exists a non-zero solution for the following system:*

$$\sum_{t=1}^{T}\left(\sum_{k=t}^{T}\gamma_k\right)(y^t - x_j^t) \geq 0, j = 1, \cdots, n,$$

$$\gamma_k \geq 0,\ k = 1,2,\dots,T,$$

$$\sum_{k=1}^{T} \gamma_k = 1.$$

Before applying this theorem to mean-risk problems we need to model the ordering of the returns of a general portfolio $w \in W$. We may do it, for example, using so-called permutations matrix $P = \{p_{i,k}\}_{i,k=1}^{T}$, that is, a 0-1 matrix that satisfies:

$$\sum_{i=1}^{T} p_{i,k} = \sum_{k=1}^{T} p_{i,k} = 1, p_{i,k} \in \{0,1\}, i,k = 1,\dots,T.$$

Then for any portfolio returns $\mathrm{x}^{\mathrm{t}}\mathrm{w}, t = 1,2,\dots n$, a permutation matrix P exists such that:

$$(X\mathbf{w})^{[t]} = \sum_{k=1}^{T} p_{t,k}\mathrm{x}^{k}\mathrm{w}\,,$$

that is, $PX\mathbf{w}$ is a vector of ordered returns from the smallest one. Using this ordering mechanism, we may combine together the mean-VaR objectives and SSD efficiency conditions and derive a new optimization problem called: mean-VaR problem with additional SSD efficiency constraints:

$$\min \gamma$$

$$\mathrm{s.t.}\ \gamma + \mathrm{M}.\,\delta_{\mathrm{t}} \geq -\mathrm{x}^{\mathrm{t}}\mathrm{w}, \mathrm{t} = 1,\dots,\mathrm{T},$$

$$\sum_{\mathrm{t}=1}^{\mathrm{T}} \delta_{\mathrm{t}} = \lfloor(1-\alpha)\mathrm{T}\rfloor,$$

$$\delta_{\mathrm{t}} \in \{0,1\}, \mathrm{t} = 1,\dots,\mathrm{T},$$

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}^{t}\mathbf{w} \geq m$$

$$y^{t} = \sum_{k=1}^{T} p_{t,k}\mathrm{x}^{k}\mathrm{w}\,,$$

$$\sum_{i=1}^{T} p_{i,k} = \sum_{k=1}^{T} p_{i,k} = 1, p_{i,k} \in \{0,1\}, i,k = 1,\dots,T,$$

$$\sum_{t=1}^{T}\left(\sum_{k=t}^{T} \gamma_k\right)\left(y^{t} - x_j^{t}\right) \geq 0, j = 1,\cdots,n,$$

$$\gamma_k \geq 0,\ k = 1,2,\dots,T,$$

$$\sum_{k=1}^{T} \gamma_k = 1,$$

where m is the minimal required mean return. Mean-variance and mean-CVaR models with SSD efficiency constraints can be derived similarly.

4. Empirical application

We follow Post and Kopa [21] in taking US stock market data from the Kenneth French library. We consider a standard set of 10 active benchmark stock portfolios as the base assets. They are formed, and annually rebalanced, based on individual stocks market capitalization of equity, each representing a decile of the cross-section of stocks in a given year. The first decile stocks (the smallest size) are called „small“ and the last decile stocks are called „large“. Furthermore, we include US Treasury bill as a riskless asset. We use data on annual excess

returns (in %) from 1982 to 2011 (30 observations). Hence we have $n = 11$ base assets and $T = 30$ scenarios. First, we present descriptive statistics of considered base assets in Table 1.

	mean	st. deviation	min	max	skewness	kurtosis
Small	8.43	26.66	-44.67	90.27	0.65	1.88
2nd decile	8.06	22.44	-37.92	60.56	0.04	-0.07
3rd decile	8.58	19.56	-34.80	49.95	-0.20	-0.21
4th decile	7.83	18.64	-30.75	47.68	-0.17	-0.15
5th decile	8.97	19.50	-36.86	45.49	-0.19	0.00
6th decile	8.75	17.01	-29.90	40.97	-0.20	-0.15
7th decile	9.32	18.44	-42.48	43.68	-0.49	0.95
8th decile	8.69	17.77	-40.89	39.67	-0.57	0.92
9th decile	8.62	17.12	-43.38	37.90	-0.82	1.67
Large	7.18	16.57	-36.56	32.47	-0.74	0.48

Table 1: Base assets 1982-2011 descriptive statistics.

Since we consider also risk free asset with zero excess returns, the possible mean return is between zero and 9.32. Therefore we consider m =0, 0.01, 0.02, ..., 9.32 and for each of these values we compute the mean-risk problem with and without additional SSD constraints. We depict the results for mean-VaR problem on Figure 1.

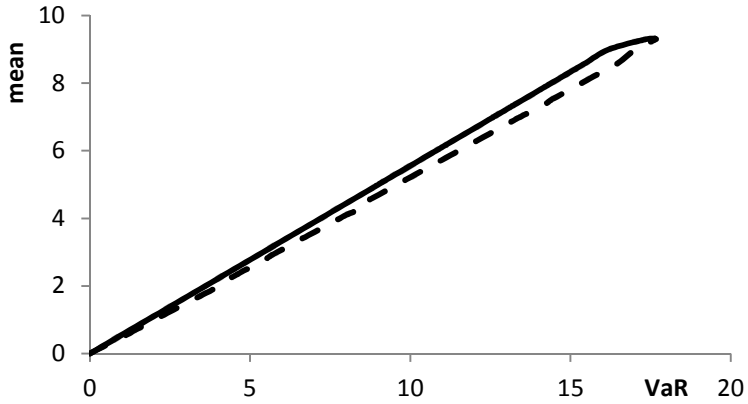


Figure 1: Mean-VaR efficiency frontiers with additional SSD efficiency constraints (dashed line) and without SSD efficincy constraints (solid line)

The figure shows that mean-VaR efficient portfolios are not SSD efficient (except of that with minimal mean return (0) or maximal mean return (9.32)). And when searching for a SSD efficient portfolio with prescribed mean return m one needs to admit a slightly higher VaR, especially for high value of parameter m.

We repeat this analysis also for mean-variance and mean-CVaR problems with (and without) SSD efficiency constraints. In these cases the difference of efficiency frontiers was very small and, moreover, for some values of parameter m the parts of efficiency frontiers with SSD efficiency constraints coincide with corresponding parts of efficiency frontiers without additionally SSD efficiency constraints. In these cases, imposing SSD efficiency constraints is harmless, because the considered mean-risk portfolios are already SSD efficient.

5. Conclusions

In this paper, we have derived and analyzed new types of optimization problems which jointly minimizes risk and maximizes mean return on the set of SSD efficient portfolios. The considered optimization problems were formulated combining three risk measures (variance, VaR and CVaR) with SSD efficiency constraints derived in Post and Kopa [21]. In the empirical application, we presented mean-risk efficiency frontiers with and without additional constraints (SSD efficiency constraints). We found that when considering VaR as a measure of risk, these frontiers differ in a subtle way. However, if variance or CVaR is considered as a measure of risk the frontiers almost coincide.

This analysis could be enriched by the optimization problems with the first order stochastic dominance (FSD), see e.g. Kopa and Post [11] or higher order stochastic dominance constraints, see e.g. Post and Kopa [21]. Alternatively, other risk measures or with other relations, for example coming from the Data envelopment analysis, see e.g. Branda and Kopa [1], [2], can be employed. Finally, one can apply the bootstrap technique as e.g. Post and Kopa [22] or consider robust versions of FSD or SSD constraints and efficiency tests see e.g. Dentcheva and Ruszczyński [4], Kopa [9] and Dupačová and Kopa [5], [6]. Unfortunately, all these modifications would either significantly increase the computational burden or lead to non-tractable formulations.

Acknowledgement The research was supported by the Czech Science Foundation (GAČR) under the project No. 402/12/0558.

References

[1] Branda, M., and Kopa, M.: DEA-risk Efficiency and Stochastic Dominance Efficiency of Stock Indices. *Czech Journal of Economics and Finance* **62**, 2 (2012), 106–124.

[2] Branda, M., and Kopa, M.: On relations between DEA-risk models and stochastic dominance efficiency tests. *Central European Journal of Operations Research* **22**, 1 (2014), 13–35.

[3] Dentcheva, D. and Ruszczyński, A.: Optimization with Stochastic Dominance Constraints. *SIAM Journal on Optimization* **14** (2003), 548–566.

[4] Dentcheva, D. and Ruszczyński, A.: Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming, Series B* **123** (2010), 85–100.

[5] Dupačová, J. and Kopa, M.: Robustness in stochastic programs with risk constraints. *Annals of Operations Research* **200**, 1 (2012), 55–74.

[6] Dupačová, J. and Kopa, M.: Robustness of optimal portfolios under risk and stochastic dominance constraints. *European Journal of Operational Research* **234**, 2 (2014), 434–441.

[7] Hadar, J. and Russell, W. R.: Rules for Ordering Uncertain Prospects. *American Economic Review* **59**, 1 (1969), 25–34.

[8] Hanoch, G. and Levy, H.: The efficiency analysis of choices involving risk. *Review of Economic Studies* **36** (1969), 335–346.

[9] Kopa, M.: Measuring of second-order stochastic dominance portfolio efficiency. *Kybernetika* **46**, 2 (2010), 488–500.

[10] Kopa, M. and Chovanec, P.: A second-order stochastic dominance portfolio efficiency measure. *Kybernetika* **44**, 2 (2008), 243–258.

[11] Kopa, M. and Post, T.: A portfolio optimality test based on the first–order stochastic dominance criterion. *Journal of Financial and Quantitative Analysis* **44**, 5 (2009), 1103–1124.

[12] Kopa, M. and Post, T.: A general test for SSD portfolio efficiency, (2014). Forthcoming in *OR Spectrum*

[13] Kuosmanen, T. Efficient diversification according to stochastic dominance criteria. *Management Science* **50**, 10 (2004), 1390–1406.

[14] Levy, H. *Stochastic dominance: Investment decision making under uncertainty*. Second edition. New York, Springer, 2006.

[15] Lizyayev, A.: Stochastic dominance efficiency analysis of diversified portfolios: classification, comparison and refinements. *Annals of Operations Research* **196**, 1 (2012), 391–410.

[16] Luedtke, J.: New Formulations for Optimization under Stochastic Dominance Constraints. *SIAM Journal on Optimization* **19**, 3 (2008), 1433–1450.

[17] Markowitz, H. M.: Portfolio Selection. *The Journal of Finance* **7**, 1 (1952), 77–91.

[18] Ogryczak, W. and Ruszczyński, A.: Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* **13** (2002), 60 – 78.

[19] Pflug, G.: Some remarks on the value–at–risk and the conditional value-at-risk. In: *Probabilistic Constrained Optimization*: *Methodology and Applications* (S. Uryasev, ed.), Kluwer Academic Publishers, Norwell MA 2000, 278–287.

[20] Post, T.: Empirical tests for stochastic dominance efficiency. *The Journal of Finance* **58** (2003), 1905–1932.

[21] Post, T. and Kopa, M.: General Linear Formulations of Stochastic Dominance Criteria. *European Journal of Operational Research* **230**, 2 (2013), 321–332.

[22] Post, T. and Kopa, M.: Aggregate Investor Preferences and Beliefs: A comment. *Journal of Empirical Finance* **23** (2013), 187–190.

[23] Quirk, J. P., and Saposnik, R.: Admissibility and Measurable Utility Functions. *Review of Economic Studies* **29** (1962), 140–146.

[24] Rockafellar, R. T and Uryasev, S.: Optimization of Conditional Value-At-Risk. *Journal of Risk* **2**, 3 (2000), 21–41.

Commodity prices and real and financial processes in the Euro area: a Bayesian SVAR approach

Maciej Kostrzewski[1], Sławomir Śmiech[2], Monika Papież[3], Marek A. Dąbrowski[4]

Abstract. This paper deals with links between global energy and non-energy commodity prices and real and financial processes in the euro area. We use monthly data spanning from 1997:1 to 2013:12 and the structural Bayesian VAR model with Sims-Zha prior specification. The analysis is performed for three sub-periods in order to capture potential changes in the reactions over time. Our main finding is that, commodity prices are indeed related to financial processes in the euro area macroeconomy: changes in the euro area interest rate have significant influence on commodity prices. There is no relation, however, between real processes in the euro area and commodity prices. Additionally, the relations between commodity prices have gradually become tighter over time.

Keywords: commodity prices, real economy, financial markets, BSVAR, Sims-Zha prior.

JEL Classification: C3, E37, E47, Q17, Q43
AMS Classification: 91G70

1 Introduction

Wide fluctuations in commodity prices have been observed in recent decades, which, according to Frankel and Rose [7], might have been caused by several factors including strong global growth (especially in China and India), easy monetary policy (low real interest rates or expected inflation), a speculative bubble and risk (geopolitical uncertainties). There are numerous articles devoted to the connections between commodity prices and financial or macroeconomic indicators. In most of them commodity prices are represented by oil prices and agricultural prices, while macroeconomic fundamentals include interest rates, inflation, exchange rates, industrial production or economic growth. Some researchers investigate the relationship between oil prices and economic growth (Cologni and Manera [5]), between oil prices and inflation rates (Chen [4]), and between oil prices and exchange rates (Śmiech and Papież [15]). Akram [1] finds evidence of a negative impact of interest rates on commodity prices, but Frankel and Rose [7] and Alquist et al. [2] do not find statistically significant relationships between real interest rates and oil prices. Tiwari [17] argues that the relationship between oil prices and German industrial production is ambiguous. Papież and Śmiech [11] analyse dependencies between the prices of crude oil and other commodities and financial investments.

The goal of this paper is to examine relations between global commodity prices and economic developments in the euro area macroeconomy in the period from 1997:1 to 2013:12. Since our intention is to include both real and financial processes we examine interdependencies between commodity prices on the one hand and economic activity (represented by industrial production) and financial conditions (represented by interest rate) in the euro area on the other hand. We also have the US dollar exchange rate in the analysis as it has been found in other studies to be related to global commodity prices (see e.g. Akram [1]). The Bayesian structural vector autoregression (BSVAR) model (see e.g. Sims and Zha [13]) is used to investigate these relations. It is quite likely that relations examined are not stable and that is why the sample is divided into three non-overlapping sub-periods.

2 Methodology

Let us consider m-dimensional structural VAR(p) model for a sample of size T (Sims and Zha [13]):

[1] Cracow University of Economics, Faculty of Management, Department of Econometrics and Operation Research, Rakowicka 27, 31-510 Cracow, Poland, e-mail: kostrzem@uek.krakow.pl.
[2] Cracow University of Economics, Faculty of Management, Department of Statistics, Rakowicka 27, 31-510 Cracow, Poland, e-mail: smiechs@uek.krakow.pl.
[3] Cracow University of Economics, Faculty of Management, Department of Statistics, Rakowicka 27, 31-510 Cracow, Poland, e-mail: papiezm@uek.krakow.pl.
[4] Cracow University of Economics, Faculty of Economics and International Relations, Department of Macroeconomics, Rakowicka 27, 31-510 Cracow, Poland, e-mail: marek.dabrowski@uek.krakow.pl.

$$y_t A_0 = y_{t-1} A_1 + \ldots + y_{t-p} A_p + c + \varepsilon_t, t = 1,\ldots,T, \tag{1}$$

where y_t is a vector of observations at time *t*, A_l is a coefficient matrix for the l^{th} lag, A_0 is a contemporaneous coefficient matrix, c is a vector of constants, ε_t is a vector of i.i.d. normal structural shocks. A distribution of ε_t conditional on y_{t-s} for $s > 0$ is normal and $E(\varepsilon_t | y_{t-s}, s > 0) = 0, E(\varepsilon_t' \varepsilon_t | y_{t-s}, s > 0) = \mathrm{I}$. Let us rewrite (1) in the matrix form so that the columns of the coefficient matrices correspond to the equations $YA_0 - XA_+ = E$. The matrix X contains the lagged *Y*'s and a column of 1's. The matrix A_+ is a matrix of the coefficients of the lagged variables and constants. The matrix E is a matrix of shocks. Define $a_0 = vec(A_0)$, $a_+ = vec(A_+)$, $A = \begin{bmatrix} A_0 \\ A_+ \end{bmatrix}$, $a = vec(A)$. The *vec* operator transforms a matrix into an vector by stacking the columns. We consider the identified model and assume that A_0 is an upper triangular nonsingular matrix.

Because of normally distributed shocks the conditional likelihood function for Y is a multivariate normal likelihood. The Sims-Zha prior is a certain modification of Litterman's structure (Sims and Zha [13], Litterman [10]). It is formed conditionally: $\pi(a) = \pi(a_0)\varphi(a_+; \mu(a_0), H(a_0))$, where $\mu(a_0)$ and $H(a_0)$ are a vector of prior mean parameters and the prior covariance for a_+, respectively. The function φ is a multivariate normal density. The prior specification is first defined for the unrestricted VAR model and then mapped into the restricted prior parameter space (see Waggoner and Zha [18]). Under some additional assumption about symmetry the posterior distribution is tractable – it can be easily sampled. The prior conditional mean is given by $E(A_+ | A_0) = \begin{bmatrix} A_0 \\ 0 \end{bmatrix}$. Sims and Zha assumed the conditional prior independent across elements of A_+ with standard deviation $\lambda_0 \lambda_1 / \sigma_j l^{\lambda_3}$ for the element corresponding to the l^{th} lag of the variable j in the equation i. The parameter λ_0 controls the tightness of beliefs on A_0, λ_1 controls tightness of the beliefs about the A_l parameters and λ_3 controls the rate at which the prior variance shrinks with increasing lag length. The hyperparameters σ_j are introduced to correct scale. The prior standard deviation of constants equals a product of $\lambda_0 \lambda_4$, where λ_4 controls tightness around the intercept. Sims and Zha specified the prior distribution of each i^{th} column of A_0 as a normal $N(0, S_i)$, where S_i is some covariance matrix. They assumed a conditional independence of A_+ but unconditional prior dependence, corresponding to A_0. The prior specification assumes the same correlation structure for the regression parameters and the structural residuals. More details about this prior specification and the hyperparameters λ_i might be found in Robertson and Tallman [12]. Posterior calculations are conducted by means of the Gibbs algorithm discussed in Waggoner and Zha [18].

3 Data

To examine relations between commodity prices on the one hand and real and financial processes in the euro area on the other hand, we use monthly data from 1997:1 to 2013:12. The data cover five series of variables. Industrial production index (IP) in the euro area describes real economy. Real 3-month interest rate in the euro area (IR) reflects financial conditional prevailing in the euro area. The data for both variables are taken from the Eurostat database. The real exchange rate of the US dollar (REX) is also included (its increase reflects the US dollar appreciation against the euro). The commodity price indices, i.e. energy price index (PENd) and non-energy commodity price index (PNENd), are based on nominal US dollars prices deflated with the US CPI. The data for these variables are taken from the databases of World Bank and Federal Reserve Bank of St. Louis. The energy price index (world trade-base weights) consists of crude oil (84.6%), natural gas (10.8%) and coal (4.6%). The non-energy commodity price index includes agriculture (64.8%), metals (31.6%), and fertilizers (3.6%). All series are expressed as indices equal 100 in 2010, seasonally adjusted and specified in natural logarithms.

The whole sample period is divided into three sub-periods: 1997:1-2002:12, 2003:1-2008:12, and 2009:1-2013:12. The division is motivated mainly by the disparate behaviour of commodity prices in these sub-periods. In the first sub-period the energy price index increased by 5.5%, while the non-energy commodity price index decreased by 29.2% (data are for indices deflated with the US CPI). In the middle sub-period, marked with large increases and decreases of commodity prices, the corresponding changes were 28.5% and 32.2%. In the last sub-period strong rebound effect was observed with 78.5% and 16.5% increases in commodity prices respectively. It

is interesting to observe that the average real interest rate was negative in the last sub-period (-0.8%) whereas it was positive in the other two sub-periods (2.0% and 0.7% on average respectively).

4 Empirical results

4.1 Time series properties of the data

A preliminary analysis of the series is carried out before estimating the main model. The standard augmented Dickey–Fuller (ADF) unit root tests for both the intercept and the trends specifications demonstrate that all variables have unit roots for each analysed sub-period. The number of lags in the test is established using AIC criterion[5]. Next, the presence of long-term relationship between integrated variables is investigated. The trace test statistic proposed by Johansen and Juselius [9] is used. If the variables are co-integrated, the VAR in first difference would not be correctly specified, and the long-term result would be very helpful in exploring the efficient parameters of short-term dynamics. According to trace test statistics and maximum eigenvalue test, there is no cointegration at 5 percent level in the first and third sub-periods. Test results demonstrate some evidence of the presence of cointegration only in the second sub-period. Trace test indicates one cointegrating equation at the 0.05 level. In contrast, maximum eigenvalue test indicates no cointegration at the 0.05 level[6]. Since the results of cointegration tests are at best ambiguous (if not suggesting the lack of cointegration), and the variables used are I(1), we use a VAR for the first differences in our five variables.

4.2 Structural impulse response analysis

We consider three BSVAR(p) models for p=1, 2, 3 under the same Sims-Zha prior specification with arbitrarily chosen hyperparameters $\lambda_0 = \lambda_1 = \lambda_3 = \lambda_4 = 1$. In other words, we choose a fairly diffuse prior distribution, which expresses the lack of prior knowledge, and shut off dummy observations (see Robertson and Tallman [12]). Selecting the most appropriate number p actually amounts to choosing the model ensuring the best data fit. The Bayesian factors point to the lag order p=3 for all sub-periods and therefore we limit our further considerations to the BSVAR(3). Results are obtained via the R package MSBVAR (Brandt and Davis [3]). The conclusions are based on 160,000 Gibbs draws, preceded by 60,000 burn-in cycles. Generating longer chains bears insignificant influence on the impulse response analysis. The impulse response analyses under the BSVAR(3) and the BSVAR(1) lead to almost the same conclusions. The main difference is that the posterior error bands are wider under the BSVAR(3). It means more uncertainty. The Gibbs sampler draws are normalized using the "DistanceMLA" method recommended in Waggoner and Zha [18]. It corresponds to the positive system shocks.

A five-variable VAR is estimated with industrial production (ΔIP), real interest rate (ΔIR), real exchange rate (ΔREX), energy price index (ΔPENd) and non-energy commodity price index (ΔPNENd). Industrial production index is set as the first variable in VAR because its adjustments to changes in other variables are assumed to rather sluggish. Since the aim of this study is to examine the response of commodity prices to real and financial processes, we place them after industrial production, real interest rate and real exchange rate. Here, we follow the ordering proposed by Akram [1] (more conventional ordering has been used by Hanson [8]). Figures 1-3 show the impulse responses to structural innovations of all our variables across sub-periods. For example, the first column of Figure 1 illustrates with a blue line the impulse response of each variable in the system to an innovation in industrial production index (output shock). Green and red lines determine 90% posterior intervals around the impulse responses based on the highest posterior density region.

In the sub-period 1997:1-2002:12 output shocks were neutral for all variables (Figure 1). The response of real interest rate, however, was positive and only marginally insignificant. Thus, the response was (almost) in line with an anti-inflationary orientation of the monetary authority: as output increases unexpectedly, the risk of inflation grows and monetary policy needs to be tightened.

Energy price index's reaction to shocks in the real interest rate was consistent with Hotteling's rule, which says that the benefit from storing a commodity should be equal to the interest rate. The benefit includes a revaluation gain and a convenience yield and is adjusted downwards by storage cost and risk premium (see Akram [1], Frankel and Rose [7] or Śmiech et al., [16]). Hotelling's rule implies that an increase in the rate of interest, given expected future commodity prices, lowers the current commodity price. Such a reaction of energy price index to the interest rate makes them similar to asset prices (Svensson [14]).

[5] Detailed test results are available from the authors upon request.

[6] Since the length of the sample is not long, and there are four series in a vector of interests, a Monte Carlo experiment is performed and the empirical critical values of trace test are determined. We find that in such case null hypothesis of no cointegration is rejected too often. For detailed information contact the authors.

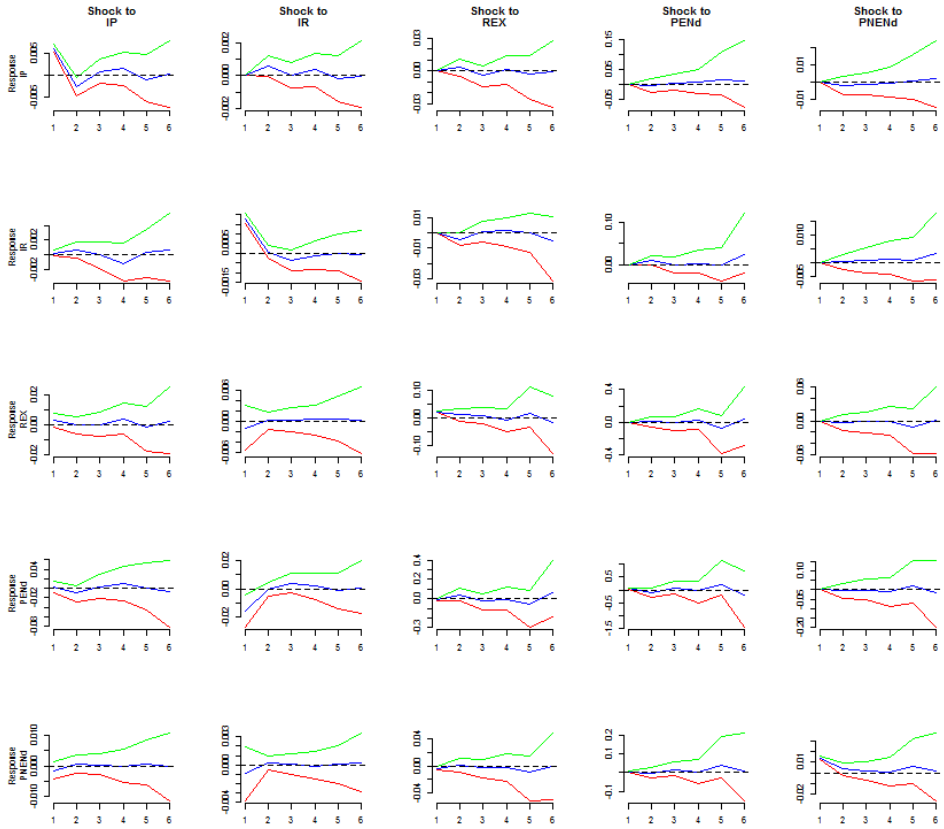


Figure 1 The impulse-responses results in sub-period 1997:1-2002:12

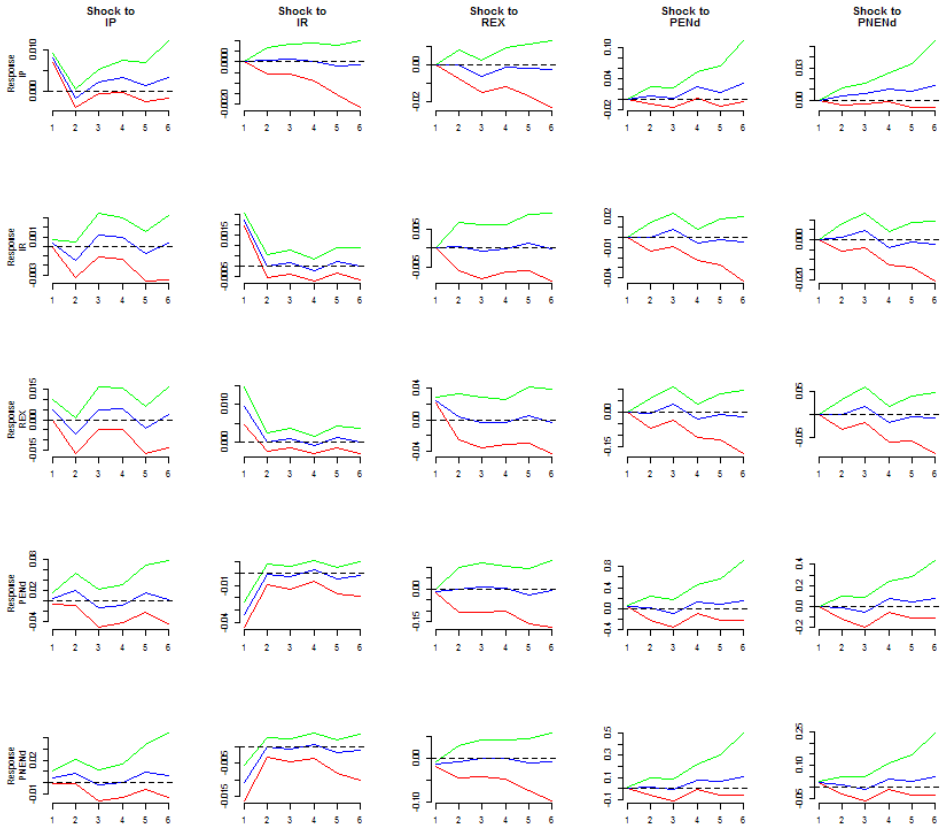


Figure 2 The impulse-responses results in sub-period 2003:1 – 2008:12

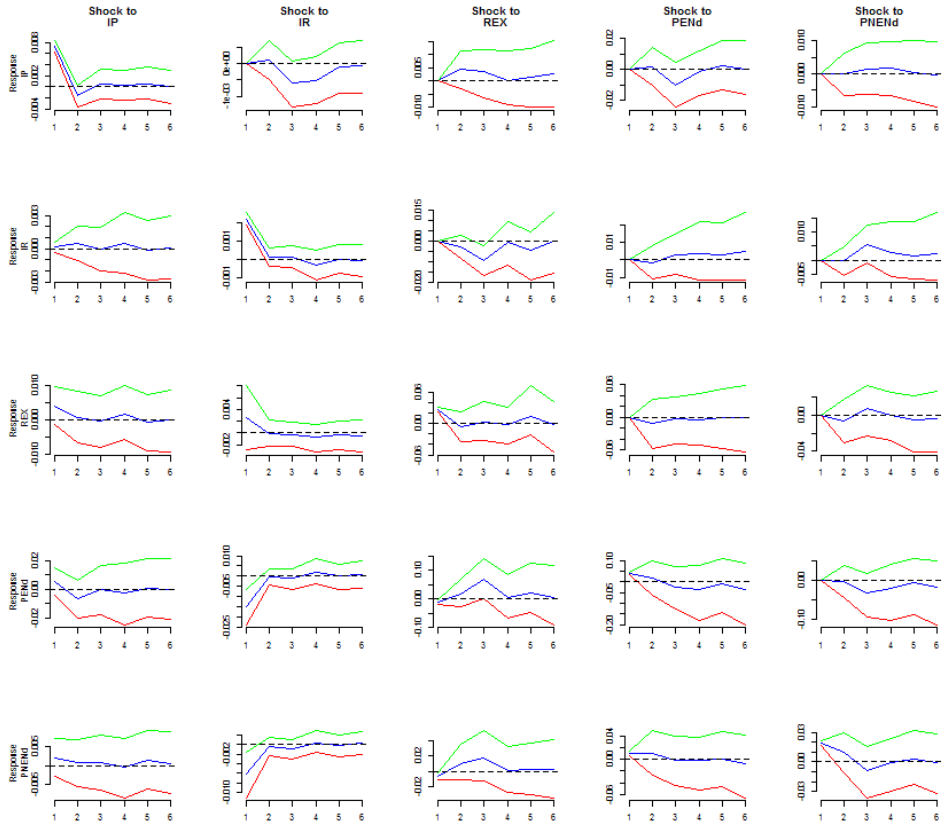


Figure 3 The impulse-responses results in sub-period 2009:1 - 2013:12

The US dollar real depreciation had a positive impact on both commodity price indices, though for energy price index at 10% level of significance only. Such reaction was found by other researchers as well, especially for the oil price (see e.g. Akram [1]). The conventional explanation is that commodity prices are quoted in US dollars, so when the dollar depreciates foreign currency commodity prices (i.e. converted into euros or pounds) are lower. Thus, the demand for commodities goes up, which results in an upward pressure on dollar price of commodities and some reversal of the initial depreciation of the US dollar.

Non-energy commodity prices respond positively to a shock in energy price index. Both commodities could be, therefore, seen as related to one another. In other words and less formally, non-energy commodity prices could not deviate too much from energy prices.

Responses in the third sub-period are almost the same as in the first sub-period (Figure 3). The only difference is that non-energy commodity price index responds to real interest rate shocks, which makes these commodities similar to energy prices.

There are more differences in the middle sub-period (Figure 2). Responses to output shocks of all variables are negligible. The importance of real shocks decreases in this sub-period and the links between both group of commodities on the one hand and the interest rate on the other hand are enhanced. It seems that commodity prices and interest rate detach from the real processes in the euro area, although not from one another. In fact, the responses of both energy and non-energy commodity price indices to interest rate shocks are twice as strong in this sub-period than in the two other sub-periods.

One more observation can be derived from Figures 1-3. The reaction of non-energy commodity price index to shocks in the energy price index has gradually increased from 0.004 to 0.009. This can be related to the developments in the biofuel market (e.g. at the share of biofuels in transport fuel consumption in the EU is estimated at 5-6%; see Demirbas [6]) and energy policy of the EU. The issue definitely requires further research.

5 Conclusion

This paper examines relations between global commodity prices and real and financial processes in the euro area macroeconomy in the period spanning from 1997:1 to 2013:12. The analysis is based on theoretical presumption that economic development in a large open economy that trades (mainly imports) commodities has non-

negligible effect on global commodity prices. We have used a Bayesian SVAR framework for three non-overlapping sub-period to allow for instability in the relations.

Our main findings are twofold. First, commodity prices are indeed related to financial processes in the euro area macroeconomy: changes in the euro area interest rate have significant influence on commodity prices. There is no relation, however, between real processes in the euro area and global commodity prices. This is especially visible in the middle sub-period, i.e. in the pre-crisis period, in which a (weak) link between interest rate and industrial production fades away as well. It seems that commodity prices and interest rate have detached from the real processes, although not from one another in this sub-period.

Second, the relation between non-energy commodity prices and energy prices has become tighter over time. This observation can be a symptom of a rising importance of biofuels in transport fuel consumption in the European Union and the EU energy policy. This, however, requires further research.

Acknowledgements

Supported by the grant No. 2012/07/B/HS4/00700 of the Polish National Science Centre.

References

[1] Akram, F.Q.: Commodity prices, interest rates and the dollar. *Energy Economics* **31** (2009), 838-851.
[2] Alquist, R., Kilian, L., and Vigfusson, R. J.: Forecasting the price of oil. No. 2011, 15. Bank of Canada Working Paper, (2011).
[3] Brandt, P., Davis, W.R.: Markov-Switching, Bayesian, Vector Autoregression Models *Multi*. Retrieved June 2, 2014, from *cran.r-project.org/web/packages/MSBVAR/MSBVAR.pdf*.
[4] Chen, S. S.: Oil price pass-through into inflation. *Energy Economics*, **31** (2009), 126-133.
[5] Cologni, A., and Manera, M.: The asymmetric effects of oil shocks on output growth: A Markov-switching analysis for the G-7 countries. *Economic Modelling*, **26** (2009), 1−29.
[6] Demirbas, A.: Competitive liquid biofuels from biomass. *Applied Energy*,**8** (2011), 17–28.
[7] Frankel, J. A., and Rose, A. K.: Determinants of agricultural and mineral commodity prices. Inflation in an Era of Relative Price Shocks. Reserve Bank of Australia (August), 2009.
[8] Hanson, M.A.: The 'price puzzle' reconsidered. *Journal of Monetary Economics*, **51** (2004), 1385–1413.
[9] Johansen, S., and Juselius, K.: Maximum likelihood estimation and inference on cointegration with applications to demand for money. *Oxford Bulletin of Economics and Statistics*, **52** (2) (1990), 169–210.
[10] Litterman, Robert B.: Forecasting with Bayesian Vector Autoregressions – Five Years of Experience. *Journal of Business and Economic Statistics* **4** (1986), 25-38.
[11] Papież, M., and Śmiech, S.: Causality in mean and variance between returns of crude oil and metal prices, agricultural prices and financial market prices. In: *Proceedings of 30th International Conference Mathematical Methods in Economics* (Ramík, J. and Stavárek, D. eds.). Karviná: Silesian University, School of Business Administration, 2012, 675-680.
[12] Robertson, J.C., Tallman, E.W.: Vector Autoregressions: Forecasting and Reality, *Economic Review* **84** (1999), 4-18.
[13] Sims, C. A., and Zha T.: Bayesian Methods for Dynamic Multivariate Models. *International Economic Review* **39** (1998), 949-366.
[14] Svensson, L.E.O.: Comment on Jeffrey Frankel, Commodity prices and monetary policy. In: Asset Prices and Monetary Policy (Campbell, J. ed.),, University of Chicago Press, Chicago, 2006.
[15] Śmiech, S., and Papież, M.: Fossil fuel prices, exchange rate, and stock market: A dynamic causality analysis on the European market. *Economics Letters*, **118(1)**, (2013), 199-202.
[16] Śmiech, S., Papież, M. and Dąbrowski, M.A.: Energy and non-energy commodity prices and the Eurozone macroeconomy: a SVAR approach, In:, Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena (Papież, M. and Śmiech, S. eds.), Foundation of the Cracow University of Economics, Cracow, (2014), 165-174.
[17] Tiwari, A.K.: Oil prices and the macroeconomy reconsideration for Germany: using continuous wavelet. *Economic Modelling*, **30** (2013), 636–642.
[18] Waggoner D. F., Zha T.: A Gibbs sampler for structural vector autoregressions, *Journal of Economic Dynamics & Control* **28** (2003), 349-366.

Uncertainty or Level of Government Expenditure - What Is More Harmful for Economy?

Igor Kotlán[1], Zuzana Machová[2], Agata Drobiszová[3]

Abstract. In these days, many developed countries are experiencing fiscal changes. It is therefore necessary to examine the various effects of fiscal policy. Current economic literature often examines the influence of fiscal variables on economic growth, but the influence of volatility of these variables is less frequently dealt with. The government expenditure uncertainty may even represent a much more problematic factor than the level of expenditure. The aim of this paper is thus to describe the effect of government expenditure on long-term economic growth with a special emphasis on the impact of volatility of government expenditure. To ensure the comprehensiveness of the model, both government expenditure and tax burden are investigated. The taxation is measured not only using a standard tax quota, but also an alternative World Tax Index.
The empirical analysis in this article is based on a dynamic panel model which uses data for OECD countries for 2000-2010. The results confirm the assumptions about the negative impact of volatility of government expenditure on growth in model with tax quota as well as in model with World Tax Index.

Keywords: Government Expenditure Volatility, Economic Growth, OECD.

JEL Classification: H20, H50, O40
AMS Classification: 62J05

1 Introduction

Nowadays, when a number of developed countries is undergoing changes in fiscal policy, it seems necessary to examine not only the level but also the volatility of government spending. If there is a high level of government expenditure (especially unproductive spending) which is almost always financed by high tax burden, it may influence economic growth negatively. A high tax burden may, in cases of the dominance of non-distortionary taxes and the invalidity of the implications of the Laffer curve, lead to higher tax revenues, which can be used to finance productive government expenditure, and sustain long-term economic growth. This partially eliminates the potential negative growth effect of high taxation. Volatility of government spending can have even more serious consequences on economic growth than the level of government spending. Because volatility of government expenditure may expand economic cycle, and effect economic cycle on economic growth is generally seen in economic theory as negative.

The aim of this paper is to use the dynamic panel model of OECD countries to describe the effect of government expenditure on long-term economic growth with a special emphasis on the impact of volatility of government expenditure.

As mentioned above, the increase in government spending is almost always financed by increased tax revenues. Therefore empirical analysis in this article includes both government expenditure and taxes. To ensure the comprehensiveness of the model taxation is measured not only using a standard tax quota, but also an alternative World Tax Index. From methodological point of view, dynamic panel model is used for OECD countries in the period 2000-2010.

2 Theoretical foundation

In an effort to achieve, insofar as possible, the best and most realistic determination of economic growth, economists try to include in their growth models a number of variables, among others those related to fiscal policy.

[1] VSB-Technical University of Ostrava, Faculty of Economics, Sokolská 33, Ostrava, Czech Republic
e-mail: igor.kotlan@vsb.cz.
[2] VSB-Technical University of Ostrava, Faculty of Economics, Sokolská 33, Ostrava, Czech Republic
e-mail: zuzana.machova@vsb.cz.
[3] VSB-Technical University of Ostrava, Faculty of Economics, Sokolská 33, Ostrava, Czech Republic
e-mail: agata.drobiszova@vsb.cz.

Interest in these types of variables has deepened as a result of the current debt crisis[4] and doubts about further EU integration. The aim of these models is therefore to confirm or refute hypotheses about the impact of fiscal variables on economic growth, especially in developed countries. Fiscal variables in particular mean the level of government expenditure and taxation, and also the size of the public sector. The most important authors dealing with this issue include Kneller, Bleaney and Gemmel [21]; recent studies include, e.g. Gemmel, Kneller and Sanz [18]. These conclude that the impact of fiscal variables on economic growth is ambiguous and depends mainly on the nature of government expenditure (productive – especially infrastructure expenditure, or non-productive – mainly social spending) and taxes (distortionary or non-distortionary), with pro-growth effects being present only in productive expenditure, if financed from non-distortionary taxes. Unproductive government expenditure and distortionary taxes have a negative impact on economic growth. Devereux and Mansoorijan [9] also point to the need to coordinate expenditure and tax policies, if fiscal policy is to positively affect growth. However, none of these studies addresses the extent to which economic growth can be influenced by the volatility of government expenditure and tax uncertainty.

Fiscal volatility represents a kind of economic risk, which is associated primarily with expectations about future developments in tax bases and tax rates, or developments in government expenditure. Taking into account this type of risk can then modify the conclusions of empirical studies dealing with the effect of government expenditure and taxation on economic growth.

There is not any consensus among economists on the relationship between government expenditure and economic growth. While Wagner [33] and his law of increasing state activity (Wagner's Law) says that economic growth leads to the growth of the public sector, including government expenditure, standard Keynesian theory understands government expenditure to be an exogenous factor that can be successfully used as a fiscal policy tool to promote economic growth. Lucas [25] and Romer [30] argue that only investments in human capital and research and development have pro-growth effects; Barro [8] also considers investments in infrastructure as having pro-growth effects. As already mentioned, it is necessary to distinguish between productive and unproductive expenditures (see e.g. Drobiszová [11]). Barro [8], Foelster and Henrekson [15], or Schaltegger and Torgler [32] add that the effect of government expenditure on growth is negative in developed, wealthy economies with a large public sector and a greater proportion of non-productive expenditure due to crowding out.

When considering the impact of changes in government expenditure on economic growth, it is also necessary to take into account its temporal nature. As indicated by Friedman [16] and Barro [7], discretionary measures in the form of changes in government expenditure can positively influence growth, but only provided that these changes are not considered permanent by economic agents. Should households and firms consider the changes permanent, it would change their consumer and investment behaviour, which could have a negative effect on economic growth.

The volatility of government expenditure may theoretically have both a positive and negative effect on economic growth. The key aspect in this case is what impact the volatility of government expenditure has on output volatility, since the influence of the economic cycle on economic growth is generally seen in economic theory as negative (Afonso and Furceri [2], Badinger [6]). By increasing or decreasing government expenditure in the Keynesian approach, it is possible to smooth out the economic cycle and positively affect mainly investment as well as consumption and economic growth; however, if fiscal policy is not in line with the cycle, if it is applied in a pro-cyclical manner, it alone may instead become the cause of output volatility (Afonso and Jalles [3]). It follows that the pressure to reduce government expenditure, and therefore its volatility, may affect growth positively, even during recessions, as it is accompanied by a reduction of discretionary measures, which may cause the cycle (Poterba [29]). As indicated by Afonso and Furceri [2], high volatility in government expenditure may also lead to the growth of interest rate volatility, which, in contradiction to the above described Keynesian approach, will negatively affect investment and economic growth.

Empirical studies confirming the positive effect of reducing discretionary measures and their volatility include Fatás and Mihov [14], Afonso and Furceri [2], Badinger [6], and Furceri [17]. The latter author, as well as previously Barro [8], adds that the effect of the volatility of government expenditure on growth varies according to the level of development of each economy, being more important in less developed countries.

Although findings about the impact of the taxes on economic growth are not the main aim of this article, this variable is included in the model to ensure the comprehensiveness of the model. Therefore it is necessary to describe the theoretical foundation how taxes influence economic growth. The size and volatility of taxation affects economic growth through its effects on individual growth variables, in particular on the level of savings and capital accumulation, or the size of the human capital and technology. In the case of taxation and its effect on growth, economists come to contradictory conclusions. Keuschnigg [20], Kotlán and Machová [23] argue that

[4] For more on this topic, see e.g. also Doman and Doman [10].

investment activities, and thus the growth, are negatively affected in particular by corporate tax, which is very often associated with decisions to place foreign direct investment, and with the taxation of dividends (e.g. Santoro and Wei [31]). Labour taxation also leads to reduced investment activities due to the pressure on corporate profits caused by a drop in labour supply (Alesina et al. [4]).

Economic theory implies that the impact of tax uncertainty on economic growth is the result of a number of different effects, and that it depends on which of them prevails. The negative impact of tax uncertainty on growth is confirmed by El-Shazly [13] and Edmiston [12]. Conversely, Niemann [28] disproves the negative impact of tax uncertainty on growth and investment, claiming that the effect is not clear, as it depends on many factors, such as risk attitude, interest rate development, the rate of return on investment, etc.

3 Dynamic panel regression: methodology and data

From a methodological point of view, the empirical analysis in this article is based on a dynamic panel model, which uses data for 34 OECD countries for 2000-2010. Real GDP per capita in purchasing power parity (GDPCAP) is therefore the dependent variable. The independent variables are then standard growth variables, namely capital accumulation, expressed as a share of investment in real GDP in purchasing power parity (INVGDP) and the accumulation of human capital as a proportion of the population in the total population aged 25-64 having completed at least upper secondary education, as classified by ISCED (HUM SC). Other explanatory variables are then fiscal variables, including the level of government expenditure, the volatility of government expenditure, taxation and the size of the tax uncertainty.

The level of government expenditure is expressed as a share of government expenditure in nominal GDP. The volatility of government expenditure is expressed by the deviation of the government's expenditure (share of GDP) in each year from an average level of government expenditure during the reporting period for each country. However, none of the variables concerns total government expenditure, but only productive expenditure, in accordance with the work of Kneller, Bleaney and Gemmell [21] and according to COFOG classification. They thus primarily include expenditure on general and public services, economic affairs, defence and education. By contrast, primarily expenditures on social affairs are excluded.

In this empirical analysis the tax burden is also included to ensure the comprehensiveness of the model. Tax uncertainty, similarly to the volatility of government expenditure, is expressed using deviations of the tax burden in each year from an average level of tax burden for the reporting period for each country. The selection of an indicator to express the tax burden is crucial both to calculate the uncertainty, and to express the level of taxation. As a standard, and also in this article, the tax burden is approximated using a tax quota, i.e. the share of tax revenues in nominal GDP; however, this brings about a whole range of negatives (see, e.g. Kotlán and Machová, [22]). As an alternative to the tax quota, the following analysis therefore also uses another indicator of the tax burden, the World Tax Index (WTI), designed for the purposes of macroeconomic comparisons and other analyses by the authors. It is a measure that combines hard data from internationally recognized sources (the OECD, the World Bank and others) with soft data obtained from an extensive survey carried out annually among tax experts from all OECD countries. The index consists of several parts: (1) Corporate Income Tax (CIT), (2) Personal Income Tax (PIT), (3) Value Added Tax (VAT), (4) Individual Property Taxes (PRO), and (5) Other Taxes on Consumption (OTC). The WTI is not only limited to tax rates, which are reflected, under certain circumstances, in the level of tax revenue and thus the tax rate; it includes other aspects associated with the tax burden, such as taxation progressivity, the administrative costs of taxation, or tax deductibility of expenses. For more details on the methodology for its construction, see Machová and Kotlán [26]. The dataset is freely available at www.worldtaxindex.com. Other data mentioned was drawn mainly from the OECD iLibrary Statistics, OECD Factbook Statistics and Penn World Table.

The regressions performed aimed to verify the hypothesis of the existence of the impact of government expenditure, and particularly volatility of government expenditure on long-term economic level and hence economic growth. In the first phase, stationarity tests were performed using a panel unit root test by Levin, Lin, Chu [24] and also by using alternative tests by Im, Pesaran, Shin [19], and Maddala, Wu [27]. The variables used were verified as stationary, but with regard to further interpretation of the results, they were expressed in logarithmic form. The lagged variables are designated (-1). Thus the model can be formulated as:

$$\begin{aligned} \mathrm{LOG(GDPCAP)}_{it} &= \alpha \mathrm{LOG(GDPCAP)}_{it-1} + \beta_1 \mathrm{LOG(INVGDP)}_{it} + \beta_2 \mathrm{LOG(HUMSC)}_{it} \\ &+ \beta_3 \mathrm{LOG(EXPROD)}_{it-1} + \beta_4 \mathrm{LOG(V\_EXPROD)}_{it-1} + \beta_5 \mathrm{LOG(TQ/WTI)}_{it-1} \\ &+ \beta_6 LOG(V\_TQ/V\_WTI)_{it-1} + u_{it} \end{aligned} \quad (1)$$

i = 1,2…34, t = 1,2…11.

The method used was panel data regression. Given the relatively small number of countries and the relatively short time series, the combination of time and cross-country data is absolutely essential. This makes the presented statistics more reliable. The software used was E-Views, version (7).

Using a robust estimator in calculating the covariance matrices ensured that the estimation results of standard deviations of parameters and hypothesis tests were correct with regard to a possible occurrence of autocorrelation and heteroscedasticity. This method is called "White Period" and it is made possible by virtue of the econometric software used.

Given that this is a dynamic panel, which includes appropriate delays of the dependent and independent variables, it cannot be reliably estimated by OLS. As the estimation technique, a generalized method of moments (GMM) was used, which included the method of instrumental variables. This method uses the Arellano-Bond estimator (Arellano, Bond [5]). The aforementioned estimation type ensures that the appropriate transformation process and using appropriate instruments eliminates the risk of endogeneity of the lagged values of the dependent variable and the independent variables with a random component. In the analyses below, the lagged values of the dependent variable were used as the instruments, with a lag of (-2). According to the Sargan test, the lagged values of the independent variables were not necessary to use.

The below model includes a lag of one period, as is usual in these types of studies (see, e.g. Acosta-Ormaechea and Yoo [1]). Alternatively, autoregression analyses with a two- and three-year delay were still carried out, with very similar results; however, with regard to the shortness of the time series, it would be impossible to reliably verify their validity from an econometric point of view.

4 Results of empirical analysis

The dynamic panel model of OECD countries for the period 2000-2010 testing both taxation proxies (TQ, WTI) was estimated as follows:

$$\begin{aligned}\mathrm{LOG(GDPCAP)}_{it} &= 0.595\,\mathrm{LOG(GDPCAP)}_{it-1} + 0.284\,\mathrm{LOG(INVGDP)}_{it} + 0.304\,\mathrm{LOG(HUMSC)}_{it} \\ &+ 0.129\,\mathrm{LOG(EXPROD)}_{it-1} - 0.001\,\mathrm{LOG(V\_EXPROD)}_{it-1} - 0.135\,\mathrm{LOG(TQ)}_{it-1} + 0.001\,LOG(V\_TQ)_{it-1}\end{aligned} \quad (2)$$

$$\begin{aligned}\mathrm{LOG(GDPCAP)}_{it} &= 0.574\,\mathrm{LOG(GDPCAP)}_{it-1} + 0.265\,\mathrm{LOG(INVGDP)}_{it} + 0.348\,\mathrm{LOG(HUMSC)}_{it} \\ &+ 0.102\,\mathrm{LOG(EXPROD)}_{it-1} - 0.001\,\mathrm{LOG(V\_EXPROD)}_{it-1} - 0.144\,\mathrm{LOG(WTI)}_{it-1} - 0.002\,LOG(V\_WTI)_{it-1}\end{aligned} \quad (3)$$

For better clarity, the results are also presented in the following *Table 1*. As already indicated, it uses lagged values of the dependent variable and other endogenous variables describing the impact of fiscal policy. This is the level of government spending and the level of taxation as well as the rate of their volatility. Given the focus of the paper, the question of particular importance is the volatility of fiscal variables. *Table 1* presents a model which uses ordinary tax rate as the tax burden approximation; and also shows the author's own tax burden index – the World Tax Index (WTI), see above.

Dependent variable	GDP per capita (PPP) LOG(GDPCAP)	GDP per capita (PPP) LOG(GDPCAP)
Number of observations	306	306
Number of instruments	34	34
J-statistics	32.97	30.854
Taxation proxy	Tax quota	World Tax Index
LOG(GDPCAP(-1))	0.595(21.263)**	0.574(19.598)**
LOG(INVGDP)	0.284(20.496)**	0.265(21.901))**
LOG(HUMSC)	0.304(4.367)**	0.348(6.529)**
LOG(EXPPROD(-1))	0.129(3.414)**	0.102(3.316)**
LOG(V_EXPPROD(-1))	-0.001(-1.930)*	-0.001(-2.162)*
LOG(TQ/WTI(-1))	-0.135(-5.351)**	-0.144(-5.037)**
LOG(V_TQ/V_WTI(-1))	0.001(2.636)**	-0.002(-6.74)**

Source: Authors' own calculations

Note: Included in parentheses are t-statistics that are adjusted for heteroskedasticity and autocorrelation; standard deviations are calculated using robust estimates; *, **, *** indicate significance levels of 10 %, 5 %, and 1 %, respectively; Arellano-Bond estimation.

Table 1 Effect of tax uncertainty and volatility of government expenditure using tax quota indicator and World Tax Index

The results presented in *Table 1* shows that substantial inertia can be seen for the GDP (GDP (-1)) indicator and a positive effect of the size of GDP in the previous period, as well as a very significant positive impact of the share of human capital (HUMSC) and investment rate (INVGDP), where the effect of percentage change in human capital is quantitatively greater than with the percentage change of physical capital.

Both models also show that there is a substantial positive impact of productive expenditure (EXPPROD(-1)) and a negative impact of taxation (TQ(-1)/WTI(-1) on economic growth. The negative impact of government spending volatility (V_EXPPROD) in both models was confirmed. Also demonstrated was the negative impact of tax uncertainty measured by indicators of effective tax rates WTI (V_WTI(-1)), but a positive impact of tax uncertainty measured by the tax quota (V_TQ(-1)). Although theoretical assumptions about the negative impact of tax uncertainty was thus not confirmed if we use the tax burden, its negative impacts are quite apparent when using this alternative indicator.

Due to the use of logarithms, the effect of these independent variables can be interpreted as the effect of their percentage change on the percentage change in output per worker. In the model with tax quota (see *Table 1,* second column) increase of productive expenditure by 10% raise the GDP per capita in the following period by about 1.29%. If we use the alternative indicator of the tax burden in the model (see *Table 1,* third column), which better describes the effective tax burden (WTI), we find that the same increasing productive expenditure is reflected in a raise of GDP per capita by 1.02 %, which is a slightly smaller pro-growth effect. From the above, one can conclude that using the alternative indicator of the tax burden leads to less positive pro-growth effects of productive expenditure.

Another issue is the volatility of productive expenditure. This variable is estimated as adversely affecting economic growth, than level of productive expenditure. There is no difference between models with tax quota and alternative indicator (WTI), in both cases the negative effect on economic growth is similar.

5 Conclusion

The current economic literature pays considerable attention to the impact of fiscal variables on economic growth. It focuses on the issue of taxation according to different types of taxes, on government expenditure and the size of the government sector, generally considering the effects of taxation to be negative, while seeing the effect of government spending, in the case of productive expenditure, as rather positive. However, studies of this type often neglect the fact that economic growth may be subject to effects other than the actual level of taxation and government expenditure. This is because fiscal variables can affect growth only due to their volatility, although their level has long been relatively low.

As a result, the aim of this paper was to use the dynamic panel model of OECD countries to describe the effect of government expenditure on long-term economic growth with a special emphasis on the impact of volatility of government expenditure.

The paper aims not only to assess the actual impact of the volatility of fiscal variables on economic growth, but in particular the use of the alternative indicator of the tax burden to calculate tax variables – the World Tax Index (WTI), which was designed by the authors and which is used in the empirical analysis in addition to the standard tax quota indicator.

The results of the empirical analysis confirm the economic theory of the positive effect of productive government spending. The positive effect of productive expenditure was demonstrated both when using the tax quota, and when using the WTI. In compound models, i.e. models using both of these tax burden indicators, assumptions about the negative impact of volatility of government expenditure were also confirmed.

Acknowledgements

This paper was financially supported within the VŠB - Technical University of Ostrava SGS grant project No. SP2014/145 “The Influence of Structure of Government Spending and Taxes on Selected Sectors and Economic Growth in the Czech Republic“.

References

[1] Acosta-Ormaechea, S., Yoo, J: Tax Composition and Growth: A Broad Cross- Country Perspective. *IMF Working Paper* (2012), WP/12/257.

[2] Afonso, A., Furceri, D.: Government Size, Composition, Volatility and Economic Growth. *European Journal of Political Economy* **26** (2010), 517-532.

[3] Afonso, A., Jalles, J. T.: Fiscal Volatility, Financial Crisis and Growth. *Applied Economic Letters* 19 (2012), 1821-1826.

[4] Alesina, A., et al.: Fiscal Policy, Profits, and Investment. *NBER Working Paper* (1999), 7207.
[5] Arellano, M., Bond, S.: Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* **58** (1991), 277–297.
[6] Badinger, H.: Cyclical Expenditure Policy, Output Volatility and Economic Growth. *Applied Economics* **44**, 7 (2012), 835-851.
[7] Barro, R. J.: Are Government Bonds Net Wealth? *Journal of Political Economy* **82**, 6 (1974): 1095-1117.
[8] Barro, R. J.: Government Spending in A Simple Model of Endogenous Growth. *The Journal of Political Economy* **98**, 5 (1990), 103–125.
[9] Devereux, M. B., Mansoorian, A.: International Fiscal-Policy Coordination and Economic-Growth. *International Economic Review* **33**, 2 (1992), 249-268.
[10] Doman, M., Doman, R.: Dynamic Linkages between Stock Markets: The Effects of Crises and Globalization. *Portuguese Economic Journal* **12**, 2(2013), 87-112.
[11] Drobiszová, A.: The Impact of Productive and Unproductive Government Expenditure on Economic Growth in OECD Countries. In: *Proceedings from the 3rd International Scientific Conference „TAXES IN THE WORLD"* (Machová, Z. and Kotlán, I., eds.). VŠB - TUO, Ostravice, 2013, 37-47.
[12] Edmiston, K. D.: Tax Uncertainty and Investment: A Cross-Country Empirical Examination. *Economic Inquiry* **42**, 3 (2001), 425-440.
[13] El-Shazly, A.: Investment Under Tax Policy Uncertainty: A Neoclassical Approach. *Public Finance Review* 37, 6 (2009), 732-749.
[14] Fatás, A., Mihov, I.: The Macroeconomic Effects of Fiscal Rules in the US States. *Journal of Public Economics* **90** (2006), 101-117.
[15] Foelster, S., Henrekson, M.: Growth Effects of Government Expenditure and Taxation in Rich Countries. *European Economic Review* **45**, 8 (2001), 1501-1520.
[16] Friedman, M.: *A Theory of the Consumption Function*. Princeton University Press, Princeton, 1957.
[17] Furceri, D.: Is Government Expenditure Volatility Harmful for Growth? A Cross-Country Analysis. *Fiscal Studies* **28**, 1 (2007), 103-120.
[18] Gemmel, N., Kneller, R., Sanz, I.: The Timing and Persistence of Fiscal Policy Impacts on Growth: Evidence from OECD Countries. *Economic Journal* **121**,550 (2011), F33-F58.
[19] Im, K. S., Pesaran, M., Shin, Y.: Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* **115** (2003), 53–74.
[20] Keuschnigg, C.: The Role of Corporate Taxation in a Large Welfare State. University of St. Gallen, Department of Economics, *Working Paper* (2009), 2009–23.
[21] Kneller, R., Bleaney, M., Gemmell, N.: Fiscal Policy and Growth: Evidence from OECD Countries. *Journal of Public Economics* **74** (1999), 171-190.
[22] Kotlán, I., Machová, Z.: The Influence of Taxation on the Economic Growth: A Failure of Tax Quota? *Politická ekonomie* **60**, 6 (2012) 743-763.
[23] Kotlán, I., Machová, Z.: The Impact of the Tax Burden on the Living Standard in OECD Countries. *Ekonomický časopis* **61**, 9 (2013), 1-12.
[24] Levin, A., Lin, C. F., Chu, C.: Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics* **108** (2002), 1–24.
[25] Lucas, R. E.: On the Mechanics of Economic Development. *Journal of Monetary Economics* **22** (1988), 3–39.
[26] Machová, Z., Kotlán, I.: World Tax Index: New Methodology for OECD Countries, 2000-2010. *DANUBE: Law and Economics Review* **4**, 2 (2013), 165-179.
[27] Maddala, G. S., Wu, S.: A Comparative Study of Unit Root Tests with Panel Data and A New Simple Test. *Oxford Bulletin of Economics and Statistics* **61** (1999), 631–652.
[28] Niemann, R.: The Impact of Tax Uncertainty on Irreversible Investment. *Review of Managerial Science* **5**, 1(2011), 1-17.
[29] Poterba, J.: State Responses to Fiscal Crisis: The Effects of Budgetary Institutions. *Journal of Political Economy* **102**, 4 (1994), 799-821.
[30] Romer, P. M.: Increasing Returns and Long-Run Growth. *Journal of Political Economy* **94**, 5 (1986), 1002–37.
[31] Santoro, M., Wei, C.: A Note on the Impact of Progressive Dividend Taxation on Investment Decisions. *Macroeconomic Dynamics* (2009), 1–11.
[32] Schaltegger, C. A., Torgler, B.: Growth Effects of Public Expenditure on the State and Local Level: Evidence from a Sample of Rich Governments. *CREMA Working Paper* (2004), 2004-16.
[33] Wagner, A.: Staat in nationalökonomischer Hinsicht. In: Conrad J et al. (eds.): *Handwörterbuch der Staatswissenschaften*. Lexis, Jena (1911), 727-739.

The using of linear programming for solving the municipal waste collection problem

Petr Kozel[1], Šárka Michalcová[2], Václav Friedrich[3]

Abstrakt. The municipal waste (MP) collection is one of the basic services which must be ensured both for the environment and for the convenience of residents. The companies which carry out the MP collection are frequently faced with the question: How to plan a route of vehicles optimally? As a tool for the decision support it is possible to use linear programming. The contribution deals with application of the mathematical model which solves some specific problems. These problems are based on practice: service of municipalities from the district Frydek–Mistek. The model of mentioned problem is presented in this paper and requisite numerical experiments with real data are described in detail. Their results are compared with the present state.

Keywords: municipal waste collection, linear programming, Eulerian path, the vehicle scheduling.

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

Collection of municipal waste is necessary both in terms of the environment and the population comfort. The main question in collection realization is: What is the optimal route planning for service vehicles that provide collection? The criterion for the optimal solution is always the total distance needed to travel during operation. Tasks that are related to the problem of municipal waste collection can be divided into two basic groups. Due to the prior attention they may be focused on:

- an operated node (vertex);
- an operated edge.

To solve tasks focusing on the node operation we search the optimal circular tour or sets of these circular tours. This type of task is often solved in connection with the collection of separated waste (paper, glass, plastic) or with the collection of containers on a specific waste (hazardous waste, etc.).

When solving tasks aimed at operated edges, the optimal move is searched. On basis of this optimal move all edges of the concerned area are operated with the minimum distance travelled. This category includes tasks of municipal waste collection in containers which are placed at each house in the operated area. These tasks can be further specified by the type of network on which the operation of sections is implemented. They are:

- an undirected network – common graph;
- a directed network – common digraph;
- a mixed network – common migraph.

The attention in this contribution is focused on the operation of edges on undirected network.

2 Formulation of problem

It is specified a network $N(V, E, l)$ which defines a common graph. This network has three attributes: $V = 1, ..., m$ represents a set of vertices, $E = 1, ..., n$ is a set of edges and l_{ij} represents a valuation of the edge ij; $i, j \in V$ in the form of the road length (in meters). The challenge is to find a route that passes through each edge

[1] Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, petr.kozel@vsb.cz.
[2] Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, sarka.michalcova@vsb.cz.
[3] Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, vaclav.friedrich@vsb.cz.

at least once with the minimum length and begins and ends at the depot. This task was first formulated by Mei-Ko-Kuan and is commonly known as The Chinese Postman Problem. [9]

The solution of this task can be found using the Eulerian path (respectively using the closed Eulerian path - Eulerian circuit) which was first formulated by Leonhard Euler. [2]

The Eulerian path in a common graph N is such a move in which each edge is visited exactly once. If the Eulerian path exists then an optimal solution also exists and corresponds to the sum valuation of all edges. The closed Eulerian path in undirected network N exists only if all vertices are even degree. Assuming that there are just two odd degrees we can search for the open Eulerian path. The closed Eulerian path corresponds to the situation in practice when the vehicle returns to the depot after the network operation. Although in the case of the open Eulerian path a vehicle operates all the edges of network, its journey ends in other than the default point. In terms of municipal waste collection is necessary to take into account also the return journey and therefore we will further deal with the closed Eulerian path.

Assuming that the conditions for existence of the Eulerian path in undirected network are met, we can search for it using the Edmonds' algorithm. [2] If the conditions for existence of the Eulerian path are not fulfilled, we must choose one of the alternative approaches and on the basis of these it is necessary to add appropriate auxiliary edges into the default network. In this way a new network is created in which the Eulerian path already exists. This addition of auxiliary edges in computational procedure corresponds (in practice) to the necessity to ride out through some sections several times – thus unproductively. The aim is to minimize this unproductively travelled distance.

3 Minimum pairing

The first of the replacement procedures, which can be used to modify the default network N, is procedure that uses a mathematical model of the minimum pairing. Mathematical model of the minimum pairing is based on model of the assignment problem. [4], [6] This problem would be described in the further text.

If conditions for the existence of the closed Eulerian path in the default network N are not satisfied, it means that there are vertices with odd degrees in the network. It can be shown that there is such an even number of these vertices. If we paired vertices of odd degree appropriately, we get the network in which all vertices are of even degree and in which we can already find the closed Eulerian path. It is necessary to select all vertices of odd degree from the default network N and to compile an auxiliary network N_1 from these vertices. The network N_1 has again three attributes: $N_1\left(V_1, E_1, d_{ij}\right)$, where $V_1 = 1, ..., m(1)$, $E_1 = 1, ..., n(1)$ and d_{ij} is a constant which represents the distance among vertices in the network N_1. Every two vertices are connected by an edge in the network N_1 - it is a complete graph (undirected graph). The evaluation of edges in the auxiliary network N_1 corresponds to the distance among the vertices in the default network N. Pairs of vertices (resp. edges) are found using a mathematical model of minimum pairing and these pairs are added into the default network N. Now it is already possible to search the closed Eulerian path.

Mathematical model of minimum pairing:

$$Min\ \frac{1}{2}\sum_{i=1}^{m(1)}\sum_{j=1}^{m(1)} d_{ij} \cdot x_{ij} \tag{1}$$

$$\sum_{i=1}^{m(1)} x_{ij} = 1 \text{ pro } j = 1, ..., m(1) \tag{2}$$

$$\sum_{j=1}^{m(1)} x_{ij} = 1 \text{ pro } i = 1, ..., m(1) \tag{3}$$

$$x_{ij} = x_{ji} \text{ pro } i,\ j = 1, ..., m(1) \tag{4}$$

$$x_{ij} \in \{0, 1\} \text{ pro } i,\ j = 1, ..., m(1) \tag{5}$$

The bivalent variable x_{ij} represents an assignment (resp. unassignment) of the vertex i to the vertex j.

The objective function (1) represents the total length of unproductively distance travelled between vertices ij in the network N_1. If the vertex i is assign to the vertex j, it must be also assigned the vertex j to the vertex i at the same time. This situation is ensured by the condition (4) in mathematical model. However, the total distance travelled is counted twice in the objective function with this condition. For this reason there is the parameter ½ in the formula of the objective function (1). The condition (2) ensures that each vertex j is assigned to exactly one vertex i. So condition (3) ensures that each vertex i is assign to exactly one vertex j. The obligatory condition (5) defines that variable x_{ij} is bivalent. An important note is that the distance matrix d_{ij} must contain prohibitive constants on the main diagonal. This will ensure that there are no loops.

Numerical experiments with this model will be presented in chapter 5.

4 Number of runs the edge

The second alternative approach to modify the default network N (in this network the conditions of the closed Eulerian path also are no fulfilled) is based on the mathematical model (6) – (9). The only information, that is in the matrix of distances l_{ij} (which corresponds to the default network N), enters into the mathematical model. This matrix contains the only values that correspond to vertices (that are connected by an edge).

Mathematical model for determination a number of runs the edge ij:

$$Min \sum_{\substack{i=1 \\ exists\ l_{ij}}}^{m} \sum_{\substack{j=1 \\ exists\ l_{ij}}}^{m} l_{ij} \cdot z_{ij} \tag{6}$$

$$z_{ij} + z_{ji} \geq 1 \text{ for } i, j = 1, ..., m \text{, where } exists\ l_{ij} \tag{7}$$

$$\sum_{\substack{j=1 \\ exists\ l_{ij}}}^{m} z_{ij} = \sum_{\substack{i=1 \\ exists\ l_{ij}}}^{m} z_{ji} \text{ for } i = 1, ..., m \text{, where } exists\ l_{ij} \tag{8}$$

$$z_{ij} \in Z^{+} \text{ for } i, j = 1, ..., m \text{, where } exists\ l_{ij} \tag{9}$$

The objective function (6) represents the total distance which is needed to be covered when finding the Eulerian path. That corresponds the total distance that will be travelled during the collection of municipal waste. The variable z_{ij} is an integer variable that symbolizes a number of runs the edge.

The condition (7) ensures that every existing edge will be included in the move at least once. The condition (8) ensures that the each visited vertex will be subsequently leaved. The obligatory condition (9) ensures that variable z_{ij} can take only non-negative integers. Numerical experiments with this model are presented in the following chapter 5.

5 Numerical experiments

This chapter is going to present numerical experiments that have been implemented using above-mentioned mathematical models. Two areas were selected from the district Frydek-Mistek for numerical experiments. It was required to design routes of operating vehicles for municipal waste collection in these areas. The experiments will be distinguished in the following text by the symbols E_1, E_2. Experiments were implemented gradually with both presented models.

Both the tested areas E_1, E_2 are of the very similar size but they have different internal complexity (number of vertices and edges in the graphs).

The table (Table 1) summarizes information characterizing the individual experiments E_1, E_2.

Experiment	Extent of the area	Number of a population	Number of vertices	Number of edges	Number of containers
E_1	8,61 km^2	1789	20	28	700
E_2	8,72 km^2	3074	167	203	1007

Table 1 The characterization of experiments E_1, E_2.

The experiments can be divided into the several steps which will be subsequently described.

These are:

- data preparation;
- realization of the calculation;
- evaluation of experiments and comparison of the results with the current state.

5.1 Data preparation

To realize the numerical experiment we used input data in the form of a matrix of direct connections. This matrix contained information on the length of roads between individual vertices.

For both tasks from practice it was necessary to prepare:

- sparse distance matrix l_{ij};
- distance matrix d_{ij};
- graphic diagrams of solved areas..

All these data were prepared using by the computational environment *Wolfram Mathematica* in which the input matrices of the direct connections were processed by the modified Floyd-Warshall algorithm. [3], [9] All auxiliary calculations in *Mathematica* were obtained with the computational time in thousandths of a second (i.e. milliseconds). A part of auxiliary calculation in *Mathematica* software was also the creation of graphic diagrams of the areas which are represented on pictures (Fig. 1 and Fig. 2). The triangle-shaped vertices are of odd degree, the rounded ones are of even degree.

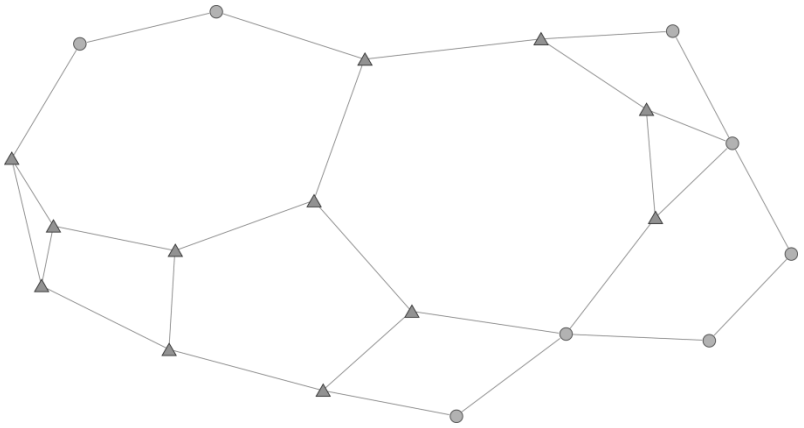


Figure 1 Graphic schema of the solved area E_1

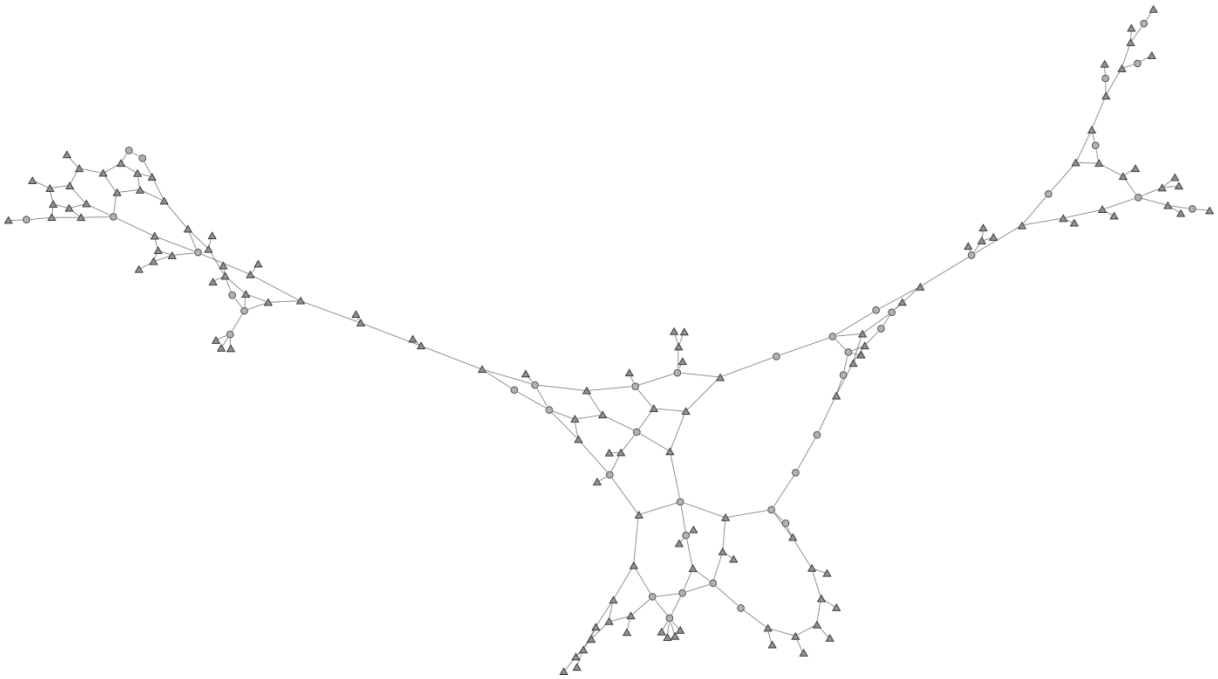


Figure 2 Graphic schema of the solved area E_2

5.2 Realization of the calculation

Numerical experiments with the above described mathematical models have been realized in computational environment *Xpresss-IVE* after the input data preparing. The computing time was under 1 second for both the tested areas.

As a first, it was tested the minimum pairing mathematical model, the input of which is information of pairs vertices pairing and the value of unproductively distance travelled. These values of both experiments E_1, E_2 are summarized at the first part of the Table 2.

Experiment	The length of unproductively distance travelled	Total distance travelled
E_1	3,76 *km*	48,75 *km*
E_2	17,84 *km*	55,5 *km*

Table 2 Numerical experiment with the minimum pairing model

To determine the total distance needed to travel during the collection of the municipal waste, the value of productively distance travelled (i.e. the length of all edges that are operated during the waste municipal collection) must be to add up to values of unproductively distance travelled. The total distance travelled is given at the second part of the Table 2.

Subsequently, the attention has been paid to the mathematical model (i.e. a number of runs the edge) which was again tested on two experiments E_1, E_2. The input of this mathematical model is information about numbers of runs the individual edges and the total distance travelled. These distances, of course, totally exactly correspond to the values of 48,75 km and 55,5 km that are presented in the Table 2. However, the results were achieved directly.

The outputs of both mathematical models can be used in the Edmond's algorithm by means of which the Eulerian path is found. This algorithm can be implemented in form of the pseudo-code in the computational environment *Xpress-IVE*. All this can be realized even within a single computational procedure. The sequence of vertices representing the Eulerian path with a minimal length is presented for experiment E_1 in the Table 3. The sequence of vertices for experiment E_2 is not presented in this article (due to a large number of values).

Experiment	The sequence of vertices representing the Eulerian path
E_1	1,2,3,5,4,2,4,6,7,5,7,9,11,12,13,14,18,19,18,20,17,18,16,15,16,20,17,12,10,9,10,8,6,8,15,1

Table 3 The sequence of vertices representing the Eulerian path

6 Conclusion

The attention was devoted to solving the task of collection of municipal waste using a linear programming in this paper. There were presented two mathematical models in this contribution. Based on these models it is possible to solve the problem assuming that the conditions of existence of the Eulerian path are not met in the network.

As a first, the mathematical model of minimum pairing was presented. This model allows determine which edges it is necessary to get through within the operation repeatedly while respecting the requirement of minimizing the total distance travelled. This mathematical model is often recommended in connection with searching of the Eulerian path. But this model has several disadvantages. The first of these is the fact that it requires more complex preparation of input data. Another one is that the output of mathematical model is further needed to process. Only then it is possible to search the Eulerian path with the help of the Edmond's algorithm.

As a second, further mathematical model was presented: number of runs the edge. The special data preparation for this model is not required. The output of model is the direct information about the total travelled distance. We can use these outputs in the Edmond's algorithm within the computational environment *Xpress-IVE* and so get the sequence of vertices corresponding to the Eulerian path with the minimal length.

Both mathematical models were tested on two practical experiments and their results can be immediately used in practice. An indicative comparison of the results (that are obtained with the help of presented procedures) with the current state (the actual distances travelled by service vehicles) is in the final Table 4.

Experiment	Total distance travelled Calculations	Total distance travelled The current state
E_1	48,75 *km*	51 *km*
E_2	55,5 *km*	74 *km*

Table 4 Comparison with the current state

References

[1] Applications of optimization with Xpress-MP, Dash Associates, Blisworth, UK, 2002, p. 240.
[2] Bondy, A., and Murty, U.: *Graph Theory.* Springer, San Francisco, 2008.
[3] Floyd, R., W.: *Algorithm 97: Shortest path.* Communications of the ACM, vol. 5, no. 6, New York, 1962.
[4] Janacek, J.: *Mathematical Programming*. EDIS, Zilina, 2003, (in Czech).
[5] Janacek, J.: *Modeling and Optimization.* Zilina, 2013, (in Slovak).
[6] Janacek, J.: *Optimization on the Ttransport Networks*. EDIS, Zilina, 2003, (in Czech).
[7] Lepikova, K.: *Waste Management of the Raškovice Municipality*. Bachelor thesis, Ostrava, 2013, (in Czech).
[8] Pelikan, J.: *The Discrete Models in Operations Research.* Professional Publishing, 2001, (in Czech).
[9] Skiena, S., and Pemmaraju, S.: *Computational Discrete Mathematics*. Cambridge University Press, New York, 2003.
[10] Wolfram, S.: *The Mathematica Book.* Cambridge University Press, New York, fourth edition, 1999.
[11] XPRESS-MP Manual "Getting Started". Dash Associates, Blisworth, UK, 2005.
[12] XPRESS-Mosel "User guide". Dash Associates, Blisworth, UK, 2005.

Application of optimization models in visualization of regional data

Jana Kramulová[1], Josef Jablonský[2], Jan Zeman[3]

Abstract. Linear assignment problem belongs among typical linear optimization problems. It can be successfully applied in the field of regional analysis. Comparisons of regions at various NUTS levels are more and more frequently used together with the rise of importance of regional policy and effort in decreasing regional disparities. When the analyses are carried out, it is necessary to present the results in a comprehensible and transparent way. One of the latest tools is a checkerplot chart which combines the depicting of size differences as well as spatial differences.
In this paper the checkerplot is constructed for 77 Czech LAU 1 regions. The first important step is to decide the size of a grid in which regional data can be presented. After that, the regions are matched with the most proper fields in the grid using the optimization assignment model. Some of the fields remain vacant. The third step consists in checking whether the spatial distribution of regions in the grid corresponds with reality. In our case, simple linear assignment model that minimizes the sum of differences between the centers of grid fields and real regional centers and model that minimizes maximum difference are compared and discussed. Finally, possible manual changes can be performed. However, application of assignment problem proves to be a useful tool for regional data visualization.

Keywords: assignment problem, minimization of maximum difference, checkerplot, regional analysis, Czech LAU 1 regions.

JEL Classification: C61
AMS Classification: 65K10, 90C05

1 Introduction

Linear assignment problem can be successfully applied in many different fields. One of them is presentation of results emerged from regional analysis. Popularity of comparisons of regions from various points of view has risen recently. If the regions included in the analysis are of a higher number (e.g. 76 in [2], or 106 in [8]), it is definitely better to present the results in the form of a chart instead of a table. Of course, there are more opportunities of charts which may be suitable for the presentation and comparison of results (for examples see e.g. [9] or [7]). One of them, checkerplot, introduced in [10], is a new graphical tool, that seems to be very useful in regional analysis while combining advantages of trellis plots (enough space for each region and good visibility) as well as of choropleth maps (spatial relation and neighborhood of regions are kept). In the procedure of constructing checkerplot, linear assignment problem is applied.

The aim of this paper is to construct checkerplot for 77 Czech LAU 1 regions, assess, if it is a useful tool for comparison of these regions and discuss, which method is more suitable and gives better and more precise results. The paper is divided into several sections. In Section 2 the checkerplot and the construction procedure are introduced. Section 3 presents linear assignment problem in two modifications – minimization of the sum of distances and minimization of the maximal distance. Section 4 shows the application in Czech conditions. Section 5 brings a brief discussion and the main conclusions.

2 The checkerplot and its construction procedure

The main advantage of checkerplot in comparison to other charts lies in the fact that it takes into consideration size differences as well as spatial differences of the regions. It means, that all regions get the same space (in a

[1] University of Economics, Prague, Department of Regional Studies, Nám. W. Churchilla 4, Praha 3, 130 67, Czech Republic, jana.kramulova@vse.cz.
[2] University of Economics, Prague, Department of Econometrics, Nám. W. Churchilla 4, Praha 3, 130 67, Czech Republic, jablon@vse.cz.
[3] University of Economics, Prague, Department of Economic Statistics, Nám. W. Churchilla 4, Praha 3, 130 67, Czech Republic, jan.zeman@vse.cz.

grid), location of each region is respected and the user is allowed to have a choice to compare different regions, under the assumption that he knows where to search for particular region.

The first task is to find a suitable grid size that would usefully represent the considered area, the regions are located in (e.g. continents, countries, regions, etc.), in the best possible ratio. Each region then corresponds to a field in a grid (usually a rectangle), i.e. the criterion of the size and shape of the region is completely suppressed. To determine the optimal distribution of regions to the grid fields, assignment problem is used as a representative of linear modelling methods. The mathematical model and further adjustments are introduced in the next section. Each region has to be determined by one particular point in space, e.g. by coordinates of the centre of each region, or by coordinates of the city that the region is represented by (when relevant).

Concerning the graphical design of resulting checkerplot, it is possible to insert symbols for each field representing the region (e.g. flags, signs, etc.) to facilitate the orientation in the graph (e.g. in case of checkerplot for EU). Regarding the chosen chart type, it is entirely the choice of the creator of checkerplot, which graph will efficiently suit the data. Almost all chart types mentioned earlier in this section may be placed in the grid fields; comprehensibility should be the deciding point. Another important issue is to ensure that all the graphs have unified scale of axis and it is, of course, assumed that all the graphs are of the same type.

In general, for the quality and sufficiently interpretable and representative checkerplot, it is important to fulfil the following points of construction procedure:

1. To find a suitable dimension of a grid and the number of fields that most closely matches the actual size of the monitored area.
2. To provide adequate and the most appropriate distribution of all regions corresponding to reality by using linear modelling techniques (assignment problem).
3. To choose the most proper type of graph (same for all grid fields) and to unify the axes of these graphs for the proper comparison of all regions.
4. Finally to add a symbol that represents the region – for greater clarity.

The first step was to determine the appropriate size of the grid, i.e. the appropriate number of the fields in both length and width, in order to get as close as possible to the real width (north to south) to length (east to west) ratio of the Czech Republic. We found this ratio of 1:0.583 and therefore our grid has the horizontal x-axis scale from 0 to 100 and the vertical y-axis scale from 0 to 58.3.

The next step is to determine the ideal number of fields. We wanted to ensure that the ratio of the number of fields in the column and row coincides or approaches as close as possible the ratio of the requested length and width of the grid. This requirement conforms almost exactly (actually the ratio is in this case 0.5833) to the number of fields 7 x 12, resulting 84 fields in the grid. While we analyze 77 LAU 1 regions in the Czech Republic, it means that in case of 7 x 12 grid just 7 fields in the grid remain vacant. The main disadvantage lies in the fact that there are not so many options how to match the regions with grid fields. Since in case of 7 x 12 grid only 7 grid fields are vacant, we decided to compute another, larger grid with more fields. Mentioned above, it is necessary to maintain the ratio of the number of fields in rows and columns corresponding to the ratio of the length and width of the Czech Republic. The 8 x 14 grid containing 112 fields fulfils this condition, with the ratio of 0.5714. In this case we obtain 35 vacant fields, which seem to be quiet many, but the occupancy rate remains high with 68.75 % of fields occupied. The different ratio also changes the coordinates on the vertical y-axis (0 to 57.14), while coordinates on the x-axis remain constant (0 to 100). After having the axis coordinates and number of fields in the row and column, it is easy to compute the field coordinates and field centers coordinates.

Subsequent task is to calculate the coordinates of the points representing the position of each district in the grid. There are two possible options how to solve this task. The first option is to determine the coordinates of points representing the centre of each district. Finding such a place as well as its at least approximately corresponding GPS coordinates in each region is a very complicated issue, since each district has a different shape. In contrast, the second option consisting in finding the GPS coordinates of the district towns is not so complicated, since these coordinates are clearly defined and traceable in many sources. For this study, we chose the second option. The GPS coordinates are re-calculated into vertical *y*-axis using equations (1)-(3)

$$Y = \frac{\text{number of minutes}}{60} + \text{number of degrees} \tag{1}$$

$$Y' = \frac{Y - SM}{NM - SM} \tag{2}$$

$$y = \text{upper boundary} \cdot Y' \tag{3}$$

where *SM* stands for the southernmost point of the Czech Republic and *NM* for the northernmost point of the Czech Republic. The re-calculation into horizontal *x*-axis is given using (4)-(6) as follows:

$$X = \frac{\text{number of minutes}}{60} + \text{number of degrees} \tag{4}$$

$$X' = 100\frac{EM - X}{EM - WM} \tag{5}$$

$$x = 100 - X' \tag{6}$$

where *EM* stands for the easternmost point of the Czech Republic and *WM* for the westernmost point of the Czech Republic.

The resulting grids with district towns are shown in Figure 1 (7 x 12 grid) and Figure 2 (8 x 14 grid). As mentioned earlier, for each district town and grid centre, the resulting coordinates *y* (3) and *x* (6) occur among 0 and upper boundary (in case of *y*) and 0 and 100 (in case of *x*). Using these coordinates the Euclidean distances are later computed. As to one field of the grid can be assigned just one region, linear modelling techniques (assignment problem) can help to provide adequate and the most appropriate distribution of all regions corresponding to reality.

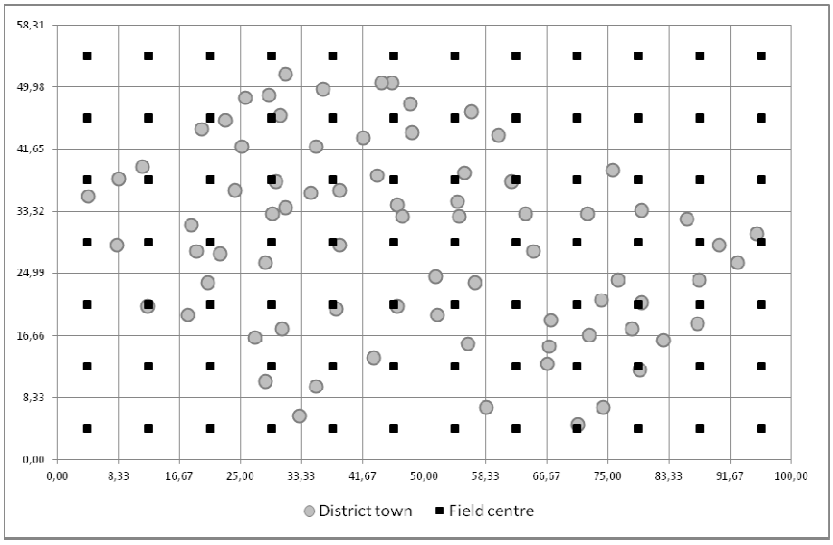


Figure 1 Grid centres' and district towns' coordinates for 7 x 12 grid

3 Assignment problem

The main aim in the problem described in the previous section is to assign district centers (towns) to the fields of the grid where each field is represented by its centre in order to minimize the sum of distances or maximum distance given by this assignment. This optimization problem can be formulated as general assignment problem which is one of typical integer linear programming problems. More information about different types of assignment problems can be found e.g. in [1]. Apart from linear assignment problem, more complicated models can be used for spatial analyses as well (e.g. hierarchical assignment model, see [5]).

Let us denote *m* the number of the rows and *n* the number of the columns of the grid and *p* the number of the district centers to be assigned to the fields. It is clear that total number of fields of the grid is $m \cdot n$ and this product must be greater than *p*. We can define binary variables x_{ij}, $i = 1, 2, \ldots, m \cdot n$, $j = 1, 2, \ldots, p$, that express assignment of the *j*-th district center to the *i*-th element of the grid. More precisely, the variables have the following meaning:

- $x_{ij} = 1$ if the j-th district center is assigned to the field of the grid in the row $[i/n] + 1$ and the column $i - n [i/n]$ where $[a]$ is the whole part of the expression *a*, and
- $x_{ij} = 0$ otherwise.

The grid on Figure 2 contains 8 rows and 14 columns and the total number of districts to be assigned is 77, i.e. the total number of variables of the assignment model is quite high – $77 \cdot 112 = 8.624$, and e.g. the value $x_{52,6} = 1$ means that the 6th district is assigned to the field in the 4th row and the 10th column.

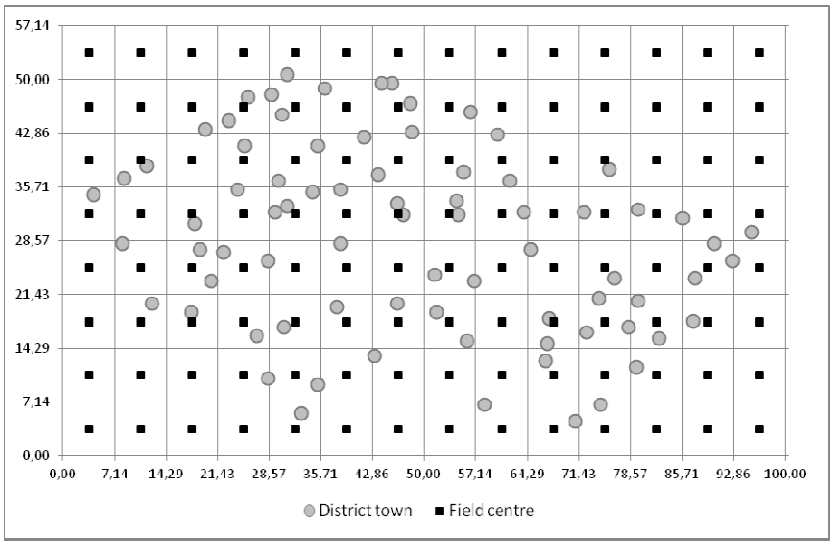


Figure 2 Grid centres' and district towns' coordinates for 8 x 14 grid

The mathematical model of the mentioned problem can be formulated as follows:

Minimise

$$z = \sum_{i=1}^{m.n} \sum_{j=1}^{p} c_{ij} x_{ij} \tag{7}$$

subject to

$$\sum_{j=1}^{p} x_{ij} \leq 1, \qquad i = 1,2,...,m \cdot n, \tag{8}$$

$$\sum_{i=1}^{m.n} x_{ij} = 1, \qquad j = 1,2,..., p. \tag{9}$$

$$x_{ij} = 0(1), \qquad i = 1,2,\ldots,m.n, j = 1,2,\ldots, p. \tag{10}$$

where c_{ij} is the Euclidean (or other) distance between the i-th field of the grid and the j-th district.

The model (7)-(10) minimises the sum of distances of the final assignment. The model can be modified in order to minimise the maximum distance of all assignments. The modified model may lead to more acceptable solutions than the first one. This result will be illustrated in the next sections of the paper. The mathematical model of assignment problem that minimises the maximum distance D of the assignments is as follows:

Minimise

$$D \tag{11}$$

subject to

$$\sum_{j=1}^{p} x_{ij} \leq 1, \qquad i = 1,2,...,m \cdot n, \tag{12}$$

$$\sum_{i=1}^{m.n} x_{ij} = 1, \qquad j = 1,2,..., p. \tag{13}$$

$$c_{ij} x_{ij} \leq D, \qquad i = 1,2,\ldots,m.n, j = 1,2,\ldots, p, \tag{14}$$

$$x_{ij} = 0(1), \qquad i = 1,2,\ldots,m.n, j = 1,2,\ldots, p. \tag{15}$$

Both presented models have $m \cdot n \cdot p$ binary variables (8.624 for 77 districts and the grid 8 x 14). The second model has one additional variable D. Model (7)-(10) has $m.n + p$ constraints (189 in our example) and model (11)-(15) has $m \cdot n(p+1) + p$ constraints (8.813 in our case). Even though it is much bigger than the first one the optimal solution using high quality solvers (CPLEX, GUROBI, XPRESS) is given in seconds.

4 Application in Czech conditions

We first computed the simple assignment problem for 7x12 grid. Upon closer examination of the distribution of districts in each field we found several of them not fully corresponding to the real location. In the 7 x 12 grid also no shape of the Czech Republic is apparent, probably due to lack of empty fields in the grid. As we see this grid as insufficient and improper, we decided to analyze if the problem with location of these districts could be solved using a larger grid.

We solved the assignment problem for the 8 x 14 grid and obtained following results. Figure 3 shows the checkerplot in 8 x 14 grid without any adjustment after solving the corresponding assignment problem.

x	x	MO	UL	**DC**	CL	LI	JN	x	x	x	x	x	x
x	x	CV	**TP**	LT	**MB**	**SM**	**JC**	TU	x	x	x	x	x
SO	KV	**RA**	LN	KD	ME	NB	HK	NA	RK	JE	BR	x	x
CH	TC	PS	BE	PZ	PY	KO	CR	**PU**	UO	SU	OP	OV	KI
x	**PM**	RO	PB	**A**	BN	KH	HB	ZR	SY	OC	PR	NJ	FM
x	DO	KT	**PJ**	PI	TA	PE	JI	BM	BK	PV	KM	VS	x
x	x	x	**ST**	**PT**	CB	JH	TR	**x**	BI	VY	UH	ZL	x
x	x	x	x	CK	x	x	x	ZN	BV	HO	x	x	x

Figure 3 Checkerplot in 8 x 14 grid without any adjustment after solving the corresponding assignment problem

The resulting distribution is now much better. We can already see the shape of the Czech Republic, as there is enough number of empty fields. However, even in this case, the analysis did not come out fully satisfactorily. Again, there are several districts that are not appropriately arranged (underlined). As we realized that this grid has already been sufficiently large, there was no need to compute another different grid. The question was whether to try to find another method that would provide a more suitable occupancy of the fields, or to use one of the previous two grids and take advantage of manual shifting of problematic districts.

Initially we decided to adjust the assignment problem in 8 x 14 grid by adding the condition of minimization of maximum difference (i.e. changing the optimizing criteria). We thought this might help to improve the results, unfortunately, the resulting grid was even worse than original grid. In the last grid we faced three empty fields inside the Czech Republic territory while not lowering the number of problematically arranged regions. This finding forced us to choose manual shifting of results in Figure 3 to improve the distribution. Such an arrangement seems to be the best available having no empty spaces inside the whole territory and maintaining the vicinity of the regions.

5 Discussion and conclusion

We would like to briefly discuss now why the necessity of manual shifts of regions to optimal distribution arises. The main reason is the selection of the deciding point to represent the region. Choosing the district town we consider to be the best option. The main advantage of this approach is definitely the simplicity of the town coordinates discovering. On the other hand the main disadvantage is probably the fact that the district town hardly ever lies in the centre of the district. Therefore when solving the assignment problem, we can face difficulties especially in the cases when the district town is located at the outskirt of the region or the region is very exten-

sive. However, we see the arrangement after solving the assignment problem as a good starting point for manual shifts, after which the final arrangement seems to be acceptable.

The aim of this paper was to assess, if the checkerplot is a useful tool for comparison of 77 Czech LAU 1 regions and alongside if it can help to improve the visualization of regional data. We constructed a grid of appropriate size so that the number of fields contributed to the optimal distribution of regions in these fields according to the reality. The grid with 14 columns and 8 rows met this assumption the most satisfactorily. However, at the beginning, we came out with a smaller grid with 12 columns and 7 rows, but this did not provide us with enough space and proper arrangement of regions in the fields. To determine borders between rows and columns as well as location of district towns we used the latitude and longitude data which we recalculated to determine the boundaries. After that midpoints of these fields (of imaginary regions) and the location of all district towns, which represent the regions, were identified.

In order to match each field in the grid with a suitable region, we used one of the integer linear programming problems, namely the assignment problem. Apart from general assignment problem we also tried to minimize the maximum distance, but this procedure did not give us better results. However, after obtaining the results from general assignment problem we were still forced to make some manual adjustments and shifts of regional areas that were not properly distributed and completely distorted reality.

It may be concluded that especially for regional analyses in which main differences are awaited in different regions (e.g. voting results) the checkerplot is a very good tool to describe (i.e. visualize) the results. Moreover, multi-criteria decision-making methods have gained a high popularity recently in the field of regional analyses, which can further increase opportunities for application of checkerplot. Most common is probably DEA (Data Envelopment Analysis) method, which can quite easily evaluate effectiveness of units, such as banks or hospitals, but is broadly used for regions as well ([8], [3] or [6]). Getting some overall value for regions (after e.g. applying composite indicator as in [4]) leads again to the problem, how to present the results. According to our opinion, checkerplot is a better choice than tables.

Acknowledgements

This research was supported by the grant no. F4/11/2013 “Optimization models in economic decision making” of the University of Economics in Prague Internal Grant Agency.

References

[1] Burkard, R. E., Çela, E.: Linear assignment problems and extensions. In: D. Z. Du and P.M. Pardalos (Eds.): Handbook of combinatorial optimization, Supplement Vol. A. 75–149. Kluwer Academic Publishers, Dordrecht, 1999.

[2] Celebioglu, F., Dall’erba, S.: Spatial Disparities across the Regions of Turkey: An Exploratory Spatial Data Analysis. *Annals of Regional Science* **45**, 2 (2010), 379–400.

[3] Fang, C., Guan, X., Lu, S., Zhou, M. and Deng, Y. Input–Output Efficiency of Urban Agglomerations in China: An Application of Data Envelopment Analysis (DEA), *Urban Studies* **50**, 13 (2013), 2766–2790.

[4] Hassan, O. A. B. Assessing the Sustainability of a Region in the Light of Composite Indicators, *Journal of Environmental Assessment Policy & Management* **10**, 1 (2008), 51–65.

[5] Horner, M. W. and O'Kelly, M. E. A Combined Cluster and Interaction Model: The Hierarchical Assignment Problem, *Geographical Analysis* 37, 3 (2005), 315–335.

[6] Rabar, D. Assessment of regional efficiency in Croatia using data envelopment analysis, *Croatian Operational Research Review* **4**, 1 (2013), 76–88.

[7] Sun, H. and Li, Z. Effectiveness of Cartogram for the Representation of Spatial Data, *The Cartographic Journal* **47**, 1 (2010), 12–21.

[8] Suzuki, S., Nijkamp, P. and Rietveld, P. Regional Efficiency Improvement by Means of Data Envelopment Analysis through Euclidean Distance Minimization Including Fixed Input Factors: An Application to Tourist Regions in Italy, *Papers in Regional Science* **90**, 1 (2011), 67–89.

[9] Templ, M., Alfons, A., Filzmoser, P., Hulliger, B. and Lussmann, D.: *Visualisation Tools*. Research Project Report (2011) WP8 – D8.2, FP7-SSH-2007-217322 AMELI (http://ameli.surveystatistics.net).

[10] Templ, M., Hulliger, B., Kowarik A., Fürst, K.: Combining geographical information and traditional plots: the checkerplot. *International Journal of Geographical Information Science* **27**, 4 (2013), 685–698.

Finding optimum fares and timetables in a monopoly rail market with heterogeneous passengers

Ondřej Krčál[1]

Abstract. This paper presents an agent-based model of passenger rail services. In the model, passengers are uniformly distributed on a circle representing the preferred departure times and passengers on each position of the circle have uniformly distributed reservation prices. The passengers choose the train with the lowest fare plus transport cost, given the sum of fare and transport cost is lower than their reservation prices. The paper provides algorithms which can be used for finding fares and timetable of a monopoly train operating company that maximizes its total profit and total welfare in realistic market settings. The efficiency of the algorithms is tested against the analytical solution of the model. The paper finds that the algorithm approximates the theoretical fare, number of trains and their departure times efficiently.

Keywords: simulation, rail, monopoly, Salop, fare, timetable.

JEL Classification: L11, L92, C63
AMS Classification: 90B20

1 Introduction

A vast majority of European long-distance passenger rail markets are served by monopoly operators who are either state-owned or selected in competitive franchising. Government agencies face difficult tasks to regulate state monopolies, or to design the franchise contracts. This paper presents a simulation tool which enables the regulators to find fares and timetables for which the monopoly operator maximizes profit, or total welfare in the market. In particular, the aim of this paper is to test the efficiency of the algorithm used in the simulation tool. For this, the paper presents a simplified version of the model which can be solved analytically using Salop model of circular city. The theoretical solution of the Salop model is compared to the results of the simulations.

Current research in transport economics uses mainly route based models of competition for studying the effects of specific timetable or pricing strategies on market outcomes. The most prominent examples are the PRAISE model developed at the Institute for Transport Studies of the University of Leeds ([1], [5], [8]) and the model used by [7]. These models do not find Nash equilibrium outcomes of competition. Instead, they study the effect of specific liberalization scenarios on profits of train companies and consumer welfare.

For finding the equilibrium price and timetable in a railway market, this paper uses agent-based technology (see [2] and [4] for examples of agent-based models of markets) which facilitates the use of the algorithm tested in this paper in more realistic situations with heterogeneous agents in future research. This paper extends the model presented in [3] in which a monopoly train operator chooses fares and timetable to maximize its profit in two respects: i) Passengers in the model [3] buy one train ticket if the fare plus transport cost is lower than their reservation price. In this paper, each passenger has a linear demand function that is decreasing in the ticket price. Hence each passenger can be interpreted as a continuum of passengers with uniformly distributed reservation prices. ii) This paper tests not only the profit-maximizing algorithm presented already in [3], but also an additional algorithm that finds fares and timetable that maximize welfare in the market.

The paper has the following structure. Section 2 introduces the model of monopolistic passenger rail market. Section 3 provides an analytical solution using the Salop model of circular city. Section 4 presents the results of simulations and compares them to the analytical solution of the model. Section 5 concludes.

2 Model

This section presents a model of a passenger railway service. The model is implemented in the modeling environment NetLogo 5.0.4 [9]. Each run of the model approximates the optimum fare and timetable in the market. In each run, the model is first initialized and then moves through a number of periods.

The initialized model contains the landscape (world), passengers and an initial set of trains. The world is a circle with a circumference equal to 240 patches (a patch may serve as a unit of distance in NetLogo). The circle

[1] Faculty of Economics and Administration, Masaryk University, department of economics, Lipová 41a, krcalo@mail.muni.cz.

represents 24 hours at a railway station. The world is populated by g identical passengers who are uniformly distributed along the line. Each passenger has a demand function $q_{ijt} = \max\{0, 1 - wh_{ijt}^2 - p_t\}$, where p_t is the fare, $w > 0$ is the per-hour weighting cost, h_{ijt} is the time passenger j waits for train i with minimum $p_t + wh_{ijt}^2$. This setting has two alternative interpretations: i) Each passenger purchases 1 unit and has the reservation price randomly drawn from uniform distribution $U[0,1] - wh_{ijt}^2$, and buys 0 units if $p_t > U[0,1] - wh_{ijt}^2$; ii) On the position of each of the g passengers, there is a continuum of passengers (which are normalized to have unit mass) with reservation prices $U[0,1] - wh_{ijt}^2$ each purchasing 1 unit of product if $p_t \leq U[0,1] - wh_{ijt}^2$, and 0 units otherwise. Finally, the model creates n_0 initial train connection. The trains are assigned an initial fare p_0 and random locations in the landscape $l_{j0} \in [0,240)$.

The initialization of the model is followed by periods which are divided in three phases: The first fare-adjusting phase takes up the first T_P periods. In this phase, the model adjusts the ticket fare given the initial number of trains and timetable. The fare-adjusting phase is supposed to eliminate the potential effect of the initial fare on simulation results. In the second phase, the model adjusts both the fare and timetable for a given number of train connections. The second phase lasts for T_S periods. The third phase, called the entry phase, consists of entry cycles. Each entry cycle has a length of E periods. The total number of entry cycles is determined in the course of the simulation.

Each period consists of two steps: 1) entry and reset, and 2) adjusting fares and timetables. First, I explain how the fares and departure times are adjusted. In the fare-adjusting phase, all trains may keep the current fare or increase or reduce the fare by a constant ε_t which is drawn out of a uniform distribution between 0 and $\varepsilon_{\max}$ in every period. In the profit-maximizing algorithm in each period $t \leq T_P$, the monopoly train operator chooses the fare $p_t + \varepsilon_t$, $p_t - \varepsilon_t$ or p_t in order to maximize its profit Π_t which is equal to the sum of profits of all its trains $\pi_{it} = p_t Q_{it}/g - F$, where Q_{it} is the sum of quantities q_{ijt} of passengers travelling by train i, g is the total number of passengers, and F is the fixed cost of the train (I divide total revenue by g because the total number of passengers is normalized to one). In the welfare-maximizing algorithm, the monopoly train operator chooses fares that maximize welfare W_t which equals to the sum of profit Π_t plus consumer surplus CS_t, which is the sum of surpluses of individual passengers equal to $0.5q_{ijt}(1 - wh_{ijt}^2 - p_t)$.

In periods $t > T_P$, the individual trains also consider leaving the departure time unchanged or changing it by a time step s_t in both directions. In each period $t > T_P$, the time step s_t is drawn from a uniform distribution between 0 and $240/(n_E\tau)$ or $240/(n_W\tau)$, where n_E or n_W are the equilibrium numbers of trains, and τ is a constant determining the maximum size of the time step. In the adjustment process in each period $t > T_P$, the train operator creates a random order of all its trains. In this order, each train chooses the fare and departure time out of the nine possible combinations of fares p_t, $p_t + \varepsilon_t$, $p_t - \varepsilon_t$ and departure times l_t, $l_t + s_t$, $l_t - s_t$ which maximizes its profit. It means that the fare may change n_t times in period t, where n_t is the total number of trains. If the profits in two or more alternatives are equal, the trains choose the departure time and fare according to the preference ordering $(l_t + s_t, p_t + \varepsilon_t) \succ (l_t + s_t, p_t - \varepsilon_t) \succ (l_t + s_t, p_t) \succ (l_t - s_t, p_t + \varepsilon_t) \succ (l_t - s_t, p_t - \varepsilon_t) \succ (l_t - s_t, p_t) \succ (l_t, p_t + \varepsilon_t) \succ (l_t, p_t - \varepsilon_t) \succ (l_t, p_t)$.

Entry and reset occur only in the entry phase. A new train connection is created in the first periods of each entry cycle, i.e. in periods $t = \{T_P + T_s + 1, T_P + T_s + E + 1, T_P + T_s + 2E + 1, \ldots\}$. The new train is assigned a random location $l_t \in [0,240)$ and the fare charged in the previous period p_{t-1}. In the last period of each entry cycle, profit or welfare (depending on whether the profit- or welfare-maximizing algorithm is used) is compared with the profit or welfare E periods back (in the last period before the latest entry). If the current profit or welfare is higher or equal ($\Pi_t \geq \Pi_{t-E}$ or $W_t \geq W_{t-E}$), the simulation continues. If it is lower, the number of trains, their location, and the fare from period $t - E$ is reset. The simulation ends when the entry fails to increase the profit or welfare (i.e. when $\Pi_t < \Pi_{t-E}$ or $W_t < W_{t-E}$) in c entry cycles in a row.

3 Analytical solution of the Salop model

The structure of the model resembles the Salop model with passengers uniformly distributed on a circle with a unit circumference served by a monopoly train operator with n trains, where the demand function of each passenger is $q_{ij} = \max\{0, 1 - wh_{ij}^2 - p\}$ (see Staněk 2014 for a detailed description and the solution of the theoretical model). The profit-maximizing monopoly distributes the trains on the circle so that the distances between all the neighboring trains are identical and all passengers are served by the operator. The optimum number of trains can be found using the following equation

$$\frac{w}{12n_E^3} = \frac{w^2}{144n_E^5} + F, \tag{1}$$

where w is the waiting cost, and n is the number of trains, and F is the fixed cost. The optimum fare equals

$$p_E = \frac{1 - \frac{w}{12{n_E}^2}}{2} \tag{2}$$

The monopoly that maximizes total welfare also distributes the trains equidistantly along the circle. The optimum number of firms is

$$n_W = \sqrt[3]{\frac{w}{6F}} \tag{3}$$

The optimum fare is $p_W = 0$. Total quantities sold by the profit- and welfare-maximizing monopoly operator are

$$\begin{aligned} q_E &= 1 - \frac{w}{12{n_E}^2} - P_E \text{ and} \\ q_W &= 1 - \frac{w}{12{n_W}^2}. \end{aligned} \tag{4}$$

The profit of the profit-maximizing operator is

$$\pi_E = p_E q_E - n_E F. \tag{5}$$

The welfare in the market under welfare maximization is given by

$$w_W = \frac{240{n_W}^4 - 40w{n_W}^2 + 3w^2}{480{n_W}^4} - n_W F. \tag{6}$$

4 Results

This section compares the results of the agent-based model with the analytical solution of the Salop model. First it presents the data generated in the simulations. Then it discusses the efficiency of the profit- and welfare-maximizing algorithms.

Both algorithms require time to find the optimum fares and timetable. The first phase has 100 periods (T_P = 100) and the second phase and the entry cycle has 50 periods ($T_S = E$ = 50). The values of the remaining parameters are as follows:

- number of passengers g = 1,000;
- fixed cost F = {0.005, 0.0075};
- equilibrium number of trains n_E or n_W = {5, 10, 15, 20};
- waiting cost is set so that the equation (1) holds, i.e. $w = 6{n_E}^2 - 6\,{n_E}^2(1 - 4\,n_E F)^{1/2}$, if the operator maximizes profit, and it is set using equation (3), i.e. $w = 6F{n_W}^3$, if the operator maximizes welfare;
- initial number of trains is n_0 = 1;
- maximum fare step $\varepsilon_{\max}$ = {0.005, 0.01};
- maximum departure-time step is $s_{\max} = 240/(\tau n_E)$ or $s_{\max} = 240/(\tau n_W)$, where n_E and n_W are optimum numbers of trains and the time-step constant is τ = {6, 12};
- number of resets in a row that ends the simulation c = 2.

I also run 4 random initializations of the model for each setting of the model using random seeds 1, 2, 3, and 4. Hence the number of runs generated for each of the algorithms is S = 128. At the end of each run, I collect the data on following variables: the number of trains N, fare P, and quantity sold by the operator Q, total profit of the monopoly operator Π, and welfare in the market W.

Comparing the measured values to the solution of the theoretical model, I find that the profit-maximizing algorithm finds the fare and timetable efficiently. The mean (standard deviation) of the difference between the simulated and theoretical number of trains $N - n_E$ equals 0.062 (0.24), of the difference in fares $P - p_E$ is $-1.3\ 10^{-4}$ ($4.17\ 10^{-4}$), of the difference in quantities $Q - q_E$ equals to $5.2\ 10^{-4}$ ($3.43\ 10^{-3}$), and of the difference in profits $\Pi - \pi_E$ is $-2.28\ 10^{-4}$ ($1.8\ 10^{-4}$). Furthermore, I use the difference between simulated and solved profits $\Pi - \pi_E$ as a measure of efficiency of the algorithm. The regression in Tab. 1 shows the effects of all exogenous variables of the model (with the exception of random seed) on the efficiency of the profit-maximizing algorithm. Fixed cost F and optimum number of trains n_E have negative and statistically significant effect on efficiency of the algorithm.

	coefficient	standard error	p-value
constant	2.93e-05	0.0001	0.78
fixed cost (F)	–0.023	0.012	0.049**
maximum fare step ($\varepsilon_{\max}$)	–0.00034	0.0059	0.954
time-step constant (τ)	4.82e-06	4.87e-06	0.324
optimum no. of trains (n_E)	–1.22e-05	2.61e-06	7.49e-06***

$S = 128$ Adj. $R^2 = 0.15$

Tab. 1: OLS estimation - dependent variable: profit difference ($\Pi - \pi_E$)

The welfare-maximizing algorithm finds the optimum number of trains somewhat less precisely compared to the profit-maximizing algorithm, but otherwise the algorithm is similarly efficient. The mean (standard deviation) of the difference between the simulated and theoretical number of trains $N - n_E$ equals –0.23 (0.44), of the difference between fares $P - p_E$ is $1.32\ 10^{-6}$ ($8.2\ 10^{-6}$), of the difference in quantities $Q - q_E$ equals to $-1.96\ 10^{-3}$ ($3.51\ 10^{-3}$), and of the difference in welfare $W - w_W$ is $-1.18\ 10^{-4}$ ($2.1\ 10^{-4}$). I use the difference in welfare $W - w_W$ as a measure of efficiency of the welfare-maximizing algorithm. Tab. 2 presents the regression results which show the effects of exogenous variables of the model on the efficiency of the algorithm. Differently from the profit-maximizing algorithm, fixed cost F and time-step constant τ increases the efficiency of the welfare-maximizing algorithm, and the optimum number of trains n_W has no effect on efficiency.

	coefficient	standard error	p-value
constant	–0.00057	0.00012	1.24e-05***
fixed cost (F)	0.036	0.014	0.0128**
maximum fare step ($\varepsilon_{\max}$)	0.0073	0.0071	0.3054
time-step constant (τ)	1.53e-05	5.94e-06	0.011**
optimum no. of trains (n_W)	2.61e-06	3.19e-06	0.4143

$S = 128$ Adj. $R^2 = 0.078$

Tab. 2: OLS estimation - dependent variable: welfare difference ($W - w_W$)

5 Conclusion

This paper presents a model of monopoly passenger rail market in which passengers differ in their preferred departure time and reservation prices. The model contains algorithms that can be used for simulating the fare and timetable that maximize profit of the monopoly train operator, or total welfare in the market. Using simplified settings for which the model can be solved analytically, the paper compares the results of the simulations with analytical solutions of the model. The paper finds that the algorithms generate results (fares, timetable, or profits and welfare) that are close to the values found analytically using the Salop model.

Acknowledgements

This paper was created as a part of the specific research project no. MUNI/A/0781/2013 at Masaryk University.

References

[1] Johnson, D., Nash, C.: Competition and the Provision of Rail Passenger Services: A Simulation Exercise. *Journal of Rail Transport Planning & Management* 2 (2012), 14–22

[2] Krčál, O.: An agent-based model of price flexing by chain-store retailers. In *Mathematical Methods in Economics* 2012. 1. vyd. Karviná: Silesian University in Opava, 2012, 461–466, ISBN 978-80-7248-779-0.

[3] Krčál, O.: Finding Equilibrium in a Monopoly Passenger Railway Market: A Comparison of Agent-Based Simulations with Theoretical Predictions. In Jedlička, P. (Ed.) *Hradec Economic Days 2014*. Hradec Králové: University of Hradec Králové, 2014, 357–363, ISBN 978-80-7435-369-7.

[4] Kvasnička, M.: Simple Agent-Based Computational Model of Market Without Intermediation. In *Mathematical Methods in Economics 2009*. Praha: Czech University of Life Sciences Prague, 2009, 204–207, ISBN 978-80-213-1963-9.
[5] Preston, J., Whelan, G., Wardman, M.: An Analysis of the Potential for On-track Competition in the British Passenger Rail Industry. *Journal of Transport Economics* **14** (1999), 77–94
[6] Staněk, R.: Market for Passenger Transport Services: Basic Framework for Welfare Analysis. In Jedlička, P. (Ed.) *Hradec Economic Days 2014*. Hradec Králové: University of Hradec Králové, 2014, 331–338, ISBN 978-80-7435-369-7.
[7] Steer Davies Gleave, EU Passenger Rail Liberalisation: Extended Impact Assessment. Report Prepared for the European Commission, 2004.
[8] Whelan, G., Johnson, D.: Modelling the Impact of Alternative Fare Structures on Train Overcrowding. *International Journal of Transport Management* **2** (2004), 51–58.
[9] Wilensky, U.: NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL, 1999.

Fuzzy Maximal Eigenvalues of Fuzzy Pairwise Comparison Matrices

Jana Krejčí[1]

Abstract. In the Analytic Hierarchy Process (AHP), the maximal eigenvalue of a pairwise comparison matrix is used to verify its consistency using the Consistency index and to obtain the weights of objects from the pairwise comparison matrix. There have been many attempts to fuzzy the AHP, among which also the fuzzification of the maximal eigenvalue of a fuzzy pairwise comparison matrix was approached. Csutora & Buckley (2001) and Wang & Chin (2006) proposed formulas to obtain the fuzzy maximal eigenvalues of fuzzy pairwise comparison matrices whose elements were triangular and trapezoidal fuzzy numbers. However, the fuzzification was not done properly. The reciprocity of fuzzy pairwise comparison matrices was not taken into account, even though it is strictly required in the AHP. We propose new formulas for computing the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix. For the sake of simplicity, only triangular fuzzy numbers are used. Afterwards, fuzzy maximal eigenvalues obtained using the new formulas are confronted with those obtained by the means of the formulas formerly proposed in the literature. Properties of the fuzzy maximal eigenvalues are described and an illustrative example is given.

Keywords: fuzzy maximal eigenvalue, fuzzy pairwise comparison matrix, reciprocity.

JEL classification: C44
AMS classification: 90C15

1 Introduction

In the Analytic Hierarchy Process, in the original method, Saaty proposed the Consistency Index CI to verify the consistency of pairwise comparison matrices and the eigenvalue method to obtain the weights of objects from pairwise comparison matrices. For the computation of the CI as well as the weights of the objects, the maximal eigenvalue of a pairwise comparison matrix is essential.

In the last few decades, the fuzzification of the AHP has become very popular among researchers since fuzzy elements can handle the vagueness of meaning of linguistic terms expressing intensities of decision makers' preferences. There have been many attempts to fuzzify the AHP, among which also the fuzzification of the eigenvalue method was proposed. In 2001, Csutora & Buckley [1] proposed formulas for obtaining the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix whose elements where either triangular or trapezoidal fuzzy numbers. After that, using the fuzzy maximal eigenvalue, they proposed an algorithm for obtaining fuzzy weights of objects from the fuzzy pairwise comparison matrix. Five years later, the fuzzification of the eigenvalue method proposed by Csutora & Buckley was revised by Wang & Chin [4]. They adopted the formulas for computing the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix with triangular fuzzy numbers and did a little modification in the formulas for obtaining the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix with trapezoidal fuzzy numbers. Afterward, they proposed a new method for obtaining the fuzzy weights of objects using the fuzzy maximal eigenvalue of the fuzzy pairwise comparison matrix. However, neither Custora & Buckley [1] nor Wang & Chin [4] took into account the requirement of the reciprocity of pairwise comparison matrices during the construction of the formulas for obtaining the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix. This is a serious flaw since the reciprocity is regarded as an essential property of pairwise comparison matrices in the AHP.

[1]University of Trento, Department of Industrial Engineering, Via Sommarive 9, 38123 Trento, Italy, jana.krejci@unitn.it

This paper focuses on the proper fuzzification of the formulas for obtaining the maximal eigenvalue of a pairwise comparison matrix. The formulas for computing the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix proposed by Csutora & Buckley [1] and by Wang & Chin [4] are reviewed, and the deficiency of the formulas regarding the violation of the reciprocity condition is pointed out. Then, improvements to the formulas are proposed in order to obtain the actual maximal eigenvalue of a fuzzy pairwise comparison matrix.

The paper is organized as follows: In Section 2, triangular fuzzy numbers and arithmetic operations with them are defined. In Section 3, a pairwise comparison matrix and a fuzzy pairwise comparison matrix are defined, and their properties are reviewed. In Section 4, formulas for obtaining the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix are dealt with. In Section 5, illustrative examples are given, and finally, in Section 6, a conclusion is formed.

2 Triangular fuzzy numbers

In this section, triangular fuzzy numbers and arithmetic operations with them are defined.

A triangular fuzzy number $\tilde{c}$ is a fuzzy number on $\mathbb{R}$ whose membership function is given as

$$\tilde{c}\,(x) = \begin{cases} \frac{x-c_l}{c_2-c_1}, & c_1 \le x \le c_2, \\ \frac{c_3-x}{c_3-c_2}, & c_2 < x \le c_3, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where c_1, c_2 and c_3 are called the lower, the middle and the upper significant values of the triangular fuzzy number $\tilde{c}$. Every triangular fuzzy number can be uniquely described by a triplet of its significant values; the notation $\tilde{c} = (c_1, c_2, c_3)$ is used in the paper hereafter. A triangular fuzzy number $\tilde{c} = (c_1, c_2, c_3)$ is said to be positive if $c_1 > 0$.

Since in the fuzzification of the AHP only positive fuzzy numbers are used, we restrict definitions of arithmetic operations only to positive fuzzy numbers. According to the simplified standard arithmetic, the sum, the product and the quotient of two positive triangular fuzzy numbers $\tilde{c} = (c_1, c_2, c_3)$ and $\tilde{d} = (d_1, d_2, d_3)$ are triangular fuzzy numbers in the form $\tilde{c} + \tilde{d} = (c_1 + d_1, c_2 + d_2, c_3 + d_3)$, $\tilde{c} * \tilde{d} = (c_1 * d_1, c_2 * d_2, c_3 * d_3)$ and $\tilde{c}/\tilde{d} = (c_1/d_3, c_2/d_2, c_3/d_1)$. The reciprocal of a triangular fuzzy number $\tilde{c}$ is a triangular fuzzy number given in the form $\frac{1}{\tilde{c}} = \left(\frac{1}{c_3}, \frac{1}{c_2}, \frac{1}{c_1}\right)$.

However, the concept of the standard fuzzy arithmetic (in this paper the simplified standard arithmetic defined above) can be applied only if there are no interactions between the fuzzy numbers. In case of any interaction between the fuzzy numbers, the constrained fuzzy arithmetic by Klir & Pan [2] should be applied on the arithmetic operations.

Let f be a continuous function, $f : \mathbb{R}^n \to \mathbb{R}$, let $\tilde{c}_i = (c_{i1}, c_{i2}, c_{i3}), i = 1, \ldots, n$, be triangular fuzzy numbers, and let D be a relation in $\mathbb{R}^n$ describing interactions among the variables. Then, according to the simplified constrained fuzzy arithmetic, $\tilde{c} = (c_1, c_2, c_3)$ is a triangular fuzzy number whose significant values are computed as

$$\begin{aligned} c_1 &= \min\{f(x_1, \ldots, x_n); (x_1, \ldots, x_n) \in D \cap [c_{11}, c_{13}] \times \cdots \times [c_{n1}, c_{n3}]\}, \\ c_2 &= f(c_{12}, \ldots, c_{n2}), \\ c_3 &= \max\{f(x_1, \ldots, x_n); (x_1, \ldots, x_n) \in D \cap [c_{11}, c_{13}] \times \cdots \times [c_{n1}, c_{n3}]\}. \end{aligned} \tag{2}$$

3 Pairwise comparison matrices and fuzzy pairwise comparison matrices

In this Section, a pairwise comparison matrix and a fuzzy pairwise comparison matrix are defined, and their properties are reviewed.

A pairwise comparison matrix of n objects is a square matrix $A = \{a_{ij}\}_{i,j=1}^{n}$ whose elements a_{ij} are numbers from Saaty's scale [5] when the object in the i-th row is more important than the object in the j-th column or their reciprocals when the object in the j-th column is more important than the object

in the i-th row. It follows that a pairwise comparison matrix has to be reciprocal, i.e. $a_{ji} = \frac{1}{a_{ij}}$ for $i, j = 1, \ldots, n, i \neq j$, and $a_{ii} = 1$ for $i = 1, \ldots, n$. A pairwise comparison matrix $A = \{a_{ij}\}_{i,j=1}^{n}$ is said to be *consistent* when $a_{ik} = a_{ij}a_{jk}$ holds for $i, j, k = 1, \ldots, n$.

The maximal eigenvalue λ_{MAX} of a pairwise comparison matrix A is the maximal solution to the equation $|A - \lambda I| = 0$ where $|.|$ denotes the determinant of a given matrix. For the sake of simplicity, the maximal eigenvalue of a given matrix is denoted only by λ hereafter. According to Perron-Frobenius theorem for nonnegative matrices follows that, for any nonnegative matrix $A = \{a_{ij}\}_{i,j=1}^{n}$, its maximal eigenvalue λ is nonnegative, i.e. $\lambda \geq 0$. Moreover, for any two matrices $A = \{a_{ij}\}_{i,j=1}^{n}, B = \{b_{ij}\}_{i,j=1}^{n}$ such that $a_{ij} \leq b_{ij}$ for $i, j = 1, \ldots, n$, and $a_{kl} < b_{kl}$ for $k, l \in \{1, \ldots, n\}$, the maximal eigenvalue λ_A of A is lower than the maximal eigenvalue λ_B of B, i.e. $\lambda_A < \lambda_B$. Further, as was shown by Saaty in [6], for any positive reciprocal matrix $A = \{a_{ij}\}_{i,j=1}^{n}$, its maximal eigenvalue λ is greater or equal to n, i.e. $\lambda \geq n$. The equality $\lambda = n$ holds if and only if the matrix $A = \{a_{ij}\}_{i,j=1}^{n}$ is consistent, i.e. $a_{ij}a_{jk} = a_{ik}$ for $i, j, k = 1, \ldots, n$.

A fuzzy pairwise comparison matrix of n objects is a square matrix $\widetilde{A} = \{\tilde{a}_{ij}\}_{i,j=1}^{n}$ whose elements $\tilde{a}_{ij}$ are fuzzy numbers from a given scale when the object in the i-th row is more important than the object in the j-th column or their reciprocals when the object in the j-th column is more important than the object in the i-th row. In this paper, fuzzified Saaty's scales of triangular fuzzy numbers defined in [3] are applied. Analogously as the pairwise comparison matrix of real numbers, also the fuzzy pairwise comparison matrix is reciprocal, i.e. $\tilde{a}_{ij} = \frac{1}{\tilde{a}_{ji}}$ for $i, j = 1, \ldots, n$. On the main diagonal of a fuzzy pairwise comparison matrix we compare always one object with itself. Naturally, the objects are always equally important as themselves and there is no fuzziness in these comparisons. Therefore, we require $\tilde{a}_{ii} = 1, i = 1, \ldots, n$, for all fuzzy pairwise comparison matrices.

4 Fuzzy maximal eigenvalue

In this section, the formulas for obtaining the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix formerly proposed in the literature are reviewed, deficiencies of the formulas are pointed out, and then new formulas are proposed. Subsequently, properties of the fuzzy maximal eigenvalue are discussed.

Csutora & Buckley [1] proposed formulas for obtaining $\alpha-$cuts, $\alpha \in [0, 1]$, of the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix. For the sake of simplicity, we focus here only on obtaining the significant values of the triangular fuzzy maximal eigenvalue. According to Csutora & Buckley [1], the significant values of the fuzzy maximal eigenvalue $\tilde{\lambda}^S = \left(\lambda_1^S, \lambda_2^S, \lambda_3^S\right)$ (the upper index S stands for the standard arithmetic which is applied to obtain the fuzzy maximal eigenvalue) of a fuzzy pairwise comparison matrix $\widetilde{A} = \{\tilde{a}_{ij}\}_{i,j=1}^{n}, \tilde{a}_{ij} = (a_{ij1}, a_{ij2}, a_{ij3})$, are computed as

$$\begin{aligned} \lambda_1^S &= max\left\{\lambda; |A_1 - \lambda I| = 0, A_1 = \{a_{ij1}\}_{i,j=1}^{n}\right\}, \\ \lambda_2^S &= max\left\{\lambda; |A_2 - \lambda I| = 0, A_2 = \{a_{ij2}\}_{i,j=1}^{n}\right\}, \\ \lambda_3^S &= max\left\{\lambda; |A_3 - \lambda I| = 0, A_3 = \{a_{ij3}\}_{i,j=1}^{n}\right\}. \end{aligned} \tag{3}$$

It means, the lower significant value λ_1^S of the fuzzy maximal eigenvalue $\tilde{\lambda}^S$ of a fuzzy pairwise comparison matrix $\widetilde{A}$ is computed as the maximal eigenvalue of the matrix $A_1 = \{a_{ij1}\}_{i,j=1}^{n}$, whose elements are the lower significant values of the triangular fuzzy numbers from the fuzzy pairwise comparison matrix $\widetilde{A}$. Analogously, the middle significant value λ_2^S is computed as the maximal eigenvalue of the matrix of the middle significant values of the triangular fuzzy numbers from the matrix $\widetilde{A}$, and the upper significant value λ_3^S is computed as the maximal eigenvalue of the matrix of the upper significant values of the triangular fuzzy numbers from the matrix $\widetilde{A}$.

Obviously, the matrices A_1 and A_3 are not reciprocal. Even Csutora & Buckley [1] themselves observed this fact. Nevertheless, they did not consider it to be a flaw. Not even Wang & Chin [4], who adopted the formulas (3) to compute fuzzy weights of objects from fuzzy pairwise comparison matrices, realized the flaw. However, as was discussed in [3], the reciprocity is a key property of fuzzy pairwise comparison matrices, which has to be preserved. It means, there are interactions between the elements of a fuzzy

pairwise comparison matrix which have to be taken into account during the computation. Therefore, the constrained fuzzy arithmetic (2) has to be applied on the formulas for obtaining the maximal eigenvalue of a pairwise comparison matrix.

By applying the simplified constrained fuzzy arithmetic, the formulas for obtaining the significant values of the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix become considerably more complicated; an optimization tool is necessary to solve the problem. Let us analyze the problem. To obtain the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix, we have to search through all reciprocal pairwise comparison matrices which can be generated by combining the elements lying in the closures of the supports of the fuzzy numbers in the fuzzy pairwise comparison matrix. We compute the maximal eigenvalue of each of these matrices, and then the lowest obtained maximal eigenvalue corresponds to the lower significant value of the resulting fuzzy maximal eigenvalue, and the greatest obtained maximal eigenvalue corresponds to the upper significant value of the resulting fuzzy maximal eigenvalue. The middle significant value of the resulting fuzzy maximal eigenvalue is computed as the maximal eigenvalue of the matrix of the middle significant values of the fuzzy numbers from the fuzzy pairwise comparison matrix since the middle significant values of the fuzzy numbers are the most possible ones and the matrix of these values is reciprocal. It means that the formula for obtaining the middle significant value of the fuzzy maximal eigenvalue remains the same as the formula proposed by Csutora & Buckley [1]. Formally, the formulas for obtaining the significant values λ_1^R, λ_2^R and λ_3^R of the maximal eigenvalue $\tilde{\lambda}^R$ (the upper index R stands for involving the reciprocity into the formulas for obtaining the fuzzy maximal eigenvalue) of a fuzzy pairwise comparison matrix $\widetilde{A} = \{\tilde{a}_{ij}\}_{i,j=1}^n, \tilde{a}_{ij} = (a_{ij1}, a_{ij2}, a_{ij3})$, are given in this form:

$$\begin{aligned}
\lambda_1^R = &\quad \min\left\{\max\left\{\lambda; |A - \lambda I| = 0, A = \{a_{ij}\}_{i,j=1}^n\right\}; a_{ij} \in [a_{ij1}, a_{ij3}], a_{ij} = \tfrac{1}{a_{ji}}, a_{ii} = 1, i, j = 1, \ldots, n\right\},\\
\lambda_2^R = &\quad \max\left\{\lambda; |A_2 - \lambda I| = 0, A_2 = \{a_{ij2}\}_{i,j=1}^n\right\},\\
\lambda_3^R = &\quad \max\left\{\max\left\{\lambda; |A - \lambda I| = 0, A = \{a_{ij}\}_{i,j=1}^n\right\}; a_{ij} \in [a_{ij1}, a_{ij3}], a_{ij} = \tfrac{1}{a_{ji}}, a_{ii} = 1, i, j = 1, \ldots, n\right\}.
\end{aligned} \tag{4}$$

Using the properties of the maximal eigenvalues reviewed in Section 3, we can derive some properties of the fuzzy maximal eigenvalues of fuzzy pairwise comparison matrices obtained both by the formulas (3) proposed by Csutora & Buckey [1] and by the new formulas (4).

First, the properties of the fuzzy maximal eigenvalue obtained by formulas (3) are summarized. Since a fuzzy pairwise comparison matrix $\widetilde{A} = \{\tilde{a}_{ij}\}_{i,j=1}^n, \tilde{a}_{ij} = (a_{ij1}, a_{ij2}, a_{ij3})$, is positive, i.e. $a_{ij1} > 0$ for $i, j = 1, \ldots, n$, the maximal eigenvalue of any matrix constructed from the elements from the closures of the supports of its fuzzy numbers is positive too. Further, because the inequalities $a_{ij1} < a_{ij2} < a_{ij3}$ hold for $i, j = 1, \ldots, n$, $i \neq j$, and $a_{ii1} = a_{ii2} = a_{ii3}$ for $i = 1, \ldots, n$, than clearly, for the maximal eigenvalues λ_1^S, λ_2^S and λ_3^S of the matrices $A_1 = \{a_{ij1}\}_{i,j=1}^n, A_2 = \{a_{ij2}\}_{i,j=1}^n$ and $A_3 = \{a_{ij3}\}_{i,j=1}^n$, the inequalities $\lambda_1^S < \lambda_2^S < \lambda_3^S$ hold. Since $A_2 = \{a_{ij2}\}_{i,j=1}^n$ is reciprocal, the inequality $\lambda_2^S \geq n$ holds for its maximal eigenvalue λ_2^S. The equality $\lambda_2^S = n$ occurs only if the matrix $A_2 = \{a_{ij2}\}_{i,j=1}^n$ is consistent, i.e. $a_{ij2}a_{jk2} = a_{ik2}$ for $i, j, k = 1, \ldots, n$.

The fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix obtained by the new formulas (4) has the following properties. Since all significant values of the fuzzy maximal eigenvalue $\tilde{\lambda}^R = \left(\lambda_1^R, \lambda_2^R, \lambda_3^R\right)$ are obtained from reciprocal matrices (see formulas (4)), the inequalities $\lambda_1^R \geq n, \lambda_2^R \geq n$ and $\lambda_3^R \geq n$ clearly hold. Further, because the lower significant value λ_1^R of the fuzzy maximal eigenvalue $\tilde{\lambda}^R$ is obtained as the minimum of a given function, the upper significant value λ_3^R is obtained as the maximum, and λ_2^R is the maximal eigenvalue of a particular matrix, the inequalities $\lambda_1^R \leq \lambda_2^R \leq \lambda_3^R$ hold. In total, for the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix obtained by formulas (4), the inequalities $n \leq \lambda_1^R \leq \lambda_2^R \leq \lambda_3^R$ hold. In special case, when the matrix A_2 of the middle significant values of the fuzzy numbers from the fuzzy pairwise comparison matrix $\widetilde{A}$ is consistent, the significant values of the fuzzy maximal eigenvalue $\tilde{\lambda}^R = \left(\lambda_1^R, \lambda_2^R, \lambda_3^R\right)$ of the fuzzy pairwise comparison matrix are in the form $n = \lambda_1^R = \lambda_2^R < \lambda_3^R$. In case that the matrix A_2 is not consistent but there exist elements in the closures of the supports of the fuzzy numbers from the fuzzy pairwise comparison matrix such that they form a consistent matrix, the significant values of the fuzzy maximal eigenvalue of the fuzzy pairwise comparison matrix are in the form $n = \lambda_1^R < \lambda_2^R < \lambda_3^R$.

Further, since λ_1^S is the maximal eigenvalue of the matrix $A = \{a_{ij1}\}_{i,j=1}^n$, λ_1^R is the maximal eigenvalue of a reciprocal matrix $A^* = \left\{a_{ij}^*\right\}_{i,j=1}^n, a_{ij}^* \in [a_{ij1}, a_{ij3}]$, and $a_{ij1} \leq a_{ij}^*$ for $i, j = 1, \ldots, n$, then obviously the inequality $\lambda_1^S < \lambda_1^R$ holds. Analogously, since λ_3^S is the maximal eigenvalue of the matrix $A = \{a_{ij3}\}_{i,j=1}^n$, λ_1^R is the maximal eigenvalue of a reciprocal matrix $A^* = \left\{a_{ij}^*\right\}_{i,j=1}^n, a_{ij}^* \in [a_{ij1}, a_{ij3}]$, and $a_{ij3} \geq a_{ij}^*$ for $i, j = 1, \ldots, n$, then also the inequality $\lambda_3^R < \lambda_3^S$ holds. Therefore, the support of the fuzzy maximal eigenvalue $\tilde{\lambda}^R$ of a given fuzzy pairwise comparison matrix is a proper subset of the support of the fuzzy maximal eigenvalue $\tilde{\lambda}^S$ of the fuzzy matrix, i.e. $\left(\lambda_1^R, \lambda_3^R\right) \subset \left(\lambda_1^S, \lambda_3^S\right)$.

By applying the reciprocity condition on the formulas for obtaining the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix, we eliminated all unfeasible combinations of elements from the supports of the fuzzy numbers in the fuzzy pairwise comparison matrix. Consequently, we obtained the fuzzy maximal eigenvalue $\tilde{\lambda}^R$ which is less vague than the fuzzy maximal eigenvalue $\tilde{\lambda}^S$ obtained by the formulas proposed by Csutora & Buckley. Moreover, the support of the fuzzy maximal eigenvalue $\tilde{\lambda}^R$ is a proper subset of the support of the fuzzy maximal eigenvalue $\tilde{\lambda}^S$ and represents the actual maximal eigenvalue of the fuzzy pairwise comparison matrix.

5 Illustrative example

In this section, three fuzzy pairwise comparison matrices are examined. The maximal eigenvalue of each of the matrices is computed both by the formulas (3) and by the new formulas (4), where the matrix reciprocity is taken into account, and the results are compared. For the construction of the fuzzy pairwise comparison matrices, fuzzified Saaty's scales defined in [3] are used.

First, let us assume the fuzzy pairwise comparison matrix $\widetilde{A}$ given in Table 1. The middle significant values of the triangular fuzzy numbers in the fuzzy pairwise comparison matrix form a consistent matrix, and therefore, as was mentioned in Section 4, its maximal eigenvalue is 3. The fuzzy maximal eigenvalue computed according to the formulas (3) is then $\tilde{\lambda}^S = (2.4983, 3, 3.7919)$, and the fuzzy maximal eigenvalue computed according to the formulas (4) is $\tilde{\lambda}^R = (3, 3, 3.0649)$. For better illustration, the fuzzy maximal eigenvalues are represented in Figure 1.

Now, let us assume the fuzzy pairwise comparison matrix $\widetilde{B}$ given in Table 1. The middle significant values of the triangular fuzzy numbers in the fuzzy matrix obviously do not form a consistent matrix. However, we can obtain a consistent matrix in the form

$$\begin{pmatrix} 1 & 3 & 9 \\ \frac{1}{3} & 1 & 3 \\ \frac{1}{9} & \frac{1}{3} & 1 \end{pmatrix} \tag{5}$$

by combining different significant values of the fuzzy numbers from the fuzzy pairwise comparison matrix $\widetilde{B}$. According to this observation, we know that the lower significant value of the fuzzy maximal eigenvalue of the fuzzy pairwise comparison matrix $\widetilde{B}$ obtained by the formulas (4) is 3. The fuzzy maximal eigenvalue computed according to the formulas (3) is then $\tilde{\lambda}^S = (2.3429, 3.0291, 4.3209)$, and the fuzzy maximal eigenvalue computed according to the formulas (4) is $\tilde{\lambda}^R = (3, 3.0291, 3.2948)$. Both the fuzzy maximal eigenvalues are represented in Figure 2.

Finally, let us assume the fuzzy pairwise comparison matrix $\widetilde{C}$ given in Table 1. Obviously, we can not obtain a consistent matrix by any combination of the elements from the closures of the supports of the fuzzy numbers from the fuzzy pairwise comparison matrix. Therefore, the lower significant value of the fuzzy maximal eigenvalue of the fuzzy pairwise comparison matrix obtained by the formulas (4) is greater than 3. The fuzzy maximal eigenvalue computed according to the formulas (3) is then $\tilde{\lambda}^S = (2.6794, 3.2085, 3.9458)$, and the fuzzy maximal eigenvalue computed according to the formulas (4) is $\tilde{\lambda}^R = (3.0291, 3.2085, 3.5608)$. Both the fuzzy maximal eigenvalues are represented in Figure 3.

Obviously, the fuzzy maximal eigenvalues of the fuzzy pairwise comparison matrices $\widetilde{A}, \widetilde{B}$ and $\widetilde{C}$ obtained by the formulas (4), which preserve the reciprocity of the matrices, are significantly less vague than the fuzzy maximal eigenvalues obtained by the formulas (3), which violate the reciprocity condition. Moreover, the closures of the supports of the fuzzy maximal eigenvalues obtained by the means of the formulas (4) are the proper subsets of the closures of the supports of the fuzzy maximal eigenvalues obtained by the means of the formulas (3).

$\widetilde{A}$			$\widetilde{B}$			$\widetilde{C}$		
1	$(1,2,3)$	$(7,8,9)$	1	$(1,3,5)$	$(7,9,9)$	1	$(3,5,7)$	$(7,9,9)$
$\left(\frac{1}{3},\frac{1}{2},1\right)$	1	$(3,4,5)$	$\left(\frac{1}{5},\frac{1}{3},1\right)$	1	$(3,5,7)$	$\left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right)$	1	$(5,7,9)$
$\left(\frac{1}{9},\frac{1}{8},\frac{1}{7}\right)$	$\left(\frac{1}{5},\frac{1}{4},\frac{1}{3}\right)$	1	$\left(\frac{1}{9},\frac{1}{9},\frac{1}{7}\right)$	$\left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right)$	1	$\left(\frac{1}{9},\frac{1}{9},\frac{1}{7}\right)$	$\left(\frac{1}{9},\frac{1}{7},\frac{1}{5}\right)$	1
$\tilde{\lambda}^S = (2.4983, 3, 3.7919)$			$\tilde{\lambda}^S = (2.3429, 3.0291, 4.3209)$			$\tilde{\lambda}^S = (2.6794, 3.2085, 3.9458)$		
$\tilde{\lambda}^R = (3, 3, 3.0649)$			$\tilde{\lambda}^R = (3, 3.0291, 3.2948)$			$\tilde{\lambda}^R = (3.0291, 3.2085, 3.5608)$		

Table 1 Fuzzy maximal eigenvalues of fuzzy pairwise comparison matrices

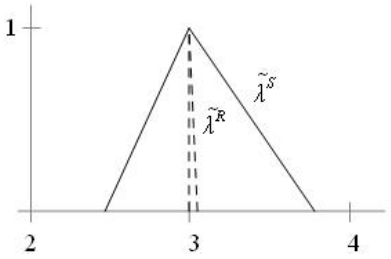


Figure 1 Fuzzy maximal eigenvalues of the fuzzy pairwise comparison matrix $\widetilde{A}$.

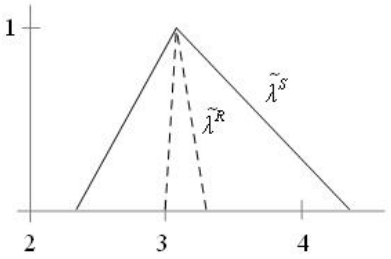


Figure 2 Fuzzy maximal eigenvalues of the fuzzy pairwise comparison matrix $\widetilde{B}$.

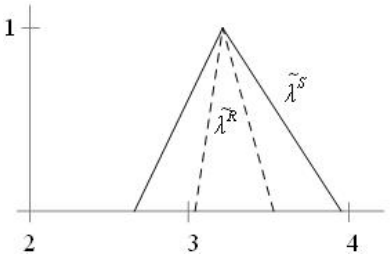


Figure 3 Fuzzy maximal eigenvalues of the fuzzy pairwise comparison matrix $\widetilde{C}$.

6 Conclusion

The proper fuzzification of the maximal eigenvalue of a pairwise comparison matrix was dealt with in this paper. The formulas for obtaining the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix of triangular fuzzy numbers proposed by Csutora & Buckley in [1] and adopted by Wang & Chin in [4] were reviewed. Deficiences of their formulas were pointed out and new formulas were introduced afterwards. The new formulas are based on the constrained fuzzy arithmetic taking into account the reciprocity condition, which is required for all pairwise comparison matrices in the AHP. Subsequently, we derived some properties of the fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix obtained both by the formulas (3) and by the new formulas (4), where the reciprocity condition is taken into account. The fuzzy maximal eigenvalue of a fuzzy pairwise comparison matrix obtained by the new formulas (4) is less vague than the fuzzy maximal eigenvalue obtained by the formulas (3) proposed by Csutora & Buckley. Moreover, it represents the actual maximal eigenvalue of the fuzzy pairwise comparison matrix since all constraints derived from the AHP and from the simplified constrained fuzzy arithmetic are taken into account in the formulas.

References

[1] Csutora, R., Buckley, J.J.: Fuzzy hierarchical analysis: the Lambda-Max method, *Fuzzy Sets and Systems* **120** (2001), 181–195.

[2] Klir, G. J., Pan, Y.: Constrained fuzzy arithmetic: Basic questions and some answers, *Soft Computing* **2** (1998), 100–108.

[3] Krejčí, J., Talašová, J.: A proper fuzzification of Saaty's scale and an improved method for computing fuzzy weights in fuzzy AHP, *Proceedings of the 31th International Conference on Mathematical Methods in Economics 2013, Jihlava* (2013), ISBN: 978-80-87035-76-4.

[4] Wang, Y.-M., Chin, K.-S.: An eigenvector method for generating normalized interval and fuzzy weights, *Applied Mathematics and Computation* **181** (2006), 1257–1275.

[5] Saaty, T. L.: *The Analytic Hierarchy Process*. McGraw Hill, New York, 1980.

[6] Saaty, T.L.: Decision-making with the AHP: Why is the principal eigenvector necessary, *European Journal of Operational Research* **145** (2003), 85–91

Commodity futures and market efficiency

Ladislav Kristoufek[1], Miloslav Vosvrda[2]

Abstract. We analyze the market efficiency of 25 commodity futures across various groups – metals, energies, soft commodities, grains and other agricultural commodities. To do so, we utilize the recently proposed Efficiency Index to find out that the most efficient among all of the analyzed commodities is heating oil, closely followed by WTI crude oil, cotton, wheat, and coffee. On the other end of the ranking scale we find live cattle and feeder cattle. The efficiency is also found to be characteristic for specific groups of commodities, with energy commodities being the most efficient and other agricultural commodities (composed mainly of livestock) the least efficient groups.

Keywords: commodities, efficiency, entropy, long-term memory, fractal dimension

JEL classification: C10, G14

1 Introduction

Efficient market hypothesis (EMH) has been a cornerstone of financial economics for decades and it has been brought to the forefront by the influential paper [9], summarizing the empirical findings based on efficient market hypotheses by [8] and [24]. Even though the actual definitions differ – the former study builds on a random walk definition and the latter on a martingale definition – the qualitative consequences are the same: the efficiency of a market originates in the impossibility of systematic control of the market, usually in the form of above-average risk-adjusted returns. [10] later subdivided the efficiency hypothesis into three forms – weak, medium and strong – which vary by the different information sets taken into consideration, and all are based on inclusion of the information sets in market prices. The weak-form EMH says that all past price movements (and associated statistics) are already reflected in the market prices. Prediction of market movements based on historical time series (technical analysis) is thus not possible for this form. The medium-form EMH states that all publicly available information is already contained in the prices, while the strong-form EMH adds all (even privately) available information. The medium form thus discards fundamental analysis and the strong form eliminates even insiders from making a profit. Evidently, a weaker form of EMH is always a subset of a stronger form. Even though EMH has been repeatedly disparaged both empirically and theoretically [3, 18], and even more so after the Global Financial Crisis broke out in 2007/2008, its validity remains an open issue, yet still it persists in standard textbooks on financial economics [7].

To the best of our knowledge, proper attention has not been given to a comparison of the efficiency of commodity markets. In this paper, we analyze futures markets for a wide range of commodities – energy, metals, and various agricultural commodities – and compare their efficiency using the Efficiency Index proposed by [15]. The paper is structured as follows. Section 2 describes

[1]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 182 08, Prague, Czech Republic, EU; Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00, Prague, Czech Republic, EU, kristouf@utia.cas.cz

[2]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 182 08, Prague, Czech Republic, EU; Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00, Prague, Czech Republic, EU, vosvrda@utia.cas.cz

the methodology in detail. Section 3 describes the analyzed dataset and gives the results. Section 4 is the conclusion.

2 Methodology

An efficient market can be defined in several ways. The main distinction has its roots back in 1965 when [8] and [24] used different definitions – a random walk and a martingale, respectively. We stick to the martingale definition of efficiency because it is less restrictive. Based on this definition, we assume that the returns of a financial asset are serially uncorrelated and with finite variance for the efficient market situation. Such a simple definition allows us to use various measures of market efficiency, which are described in this section. Eventually, we refer to the Efficiency Index which takes these statistics into consideration and it helps to rank different assets according to their efficiency while using various dynamic properties of the time series under study.

2.1 Long-term memory

Long-term memory (long-range dependence) series are characterized by values in the (even distant, in theory infinitely distant) past influencing the present and future values. These processes are typically described with the long-term memory parameter H (Hurst exponent) which ranges between $0 \leq H < 1$ for stationary invertible processes. The midpoint, $H = 0.5$, holds for uncorrelated (or in general short-term memory) processes, i.e., processes of the efficient market. For $H > 0.5$, the processes are positively correlated with long-term memory and are usually referred to as persistent. These processes systematically follow local trends while still remaining stationary. For $H < 0.5$, we have long-term memory processes with negative correlations – anti-persistent processes. Such processes switch direction more often than a random process does.

More formally, the long-term memory processes are defined in both time and frequency domains. In the time domain, they are connected to a power-law decaying auto-correlation function. For the auto-correlation function $\rho(k)$ with time lag k, the decay is characterized as $\rho(k) \propto k^{2H-2}$ for $k \to +\infty$. In the frequency domain, the spectrum $f(\lambda)$ with frequency λ of the long-range dependent process diverges at the origin so that $f(\lambda) \propto \lambda^{1-2H}$ for $\lambda \to 0+$. These definitions further lead to non-summable auto-correlations and diverging covariance of partial sums of the process for the persistent series. Such properties are used in various estimators of parameter H. For comparison of both time and frequency domain estimators, see [2, 25, 26, 23, 11, 5, 4, 1, 27]. Out of theses estimators, we opt for the local Whittle and GPH estimators (explained later in this Section) which are suitable for short time series with a possible weak short-term memory [25, 26], which can easily bring a bias into the time domain estimators [27, 14]. Moreover, these estimators have well-defined asymptotic properties – they are consistent and asymptotically normal estimators [11, 2, 23, 19].

2.2 Fractal dimension

Contrary to the long-term memory, which can be seen as a characteristic of global dependence and correlation structure, the fractal dimension D can be taken as a measure of local memory of the series as it is a measure of roughness of the series [15]. As specific parts of a series can be different with regard to their roughness or smoothness, the series can be locally serially correlated even though, on a global level, the correlations might vanish and are not necessarily observable or detectable.

For a univariate series, the fractal dimension ranges between $1 < D \leq 2$. For self-similar processes, the fractal dimension is tightly connected to the Hurst exponent (long-term memory) of the series so that $D = 2 - H$. In economic terms, this can be understood as a perfect transmission

of a local behavior (fractal dimension) to a global behavior (long-term memory). However, this relationship usually does not hold perfectly for the financial series so that both D and H give different insights into the dynamics of the series, making it worth studying them separately [15, 16].

In general, $D = 1.5$ holds for an uncorrelated series with no local trending or no local anti-correlations and thus it is also a value of D for the efficient market. For a low fractal dimension, $D < 1.5$, the roughness of the series is lower than for an uncorrelated process; hence we observe local trending and the series is said to be locally persistent. Conversely, a high fractal dimension, $D > 1.5$, characterizes a series rougher than the uncorrelated one, which is connected to local anti-persistence, i.e., such series are negatively auto-correlated locally. For purposes of the Efficiency Index introduced later in this section, we utilize Hall-Wood and Genton estimators [12, 13]. Statistical properties of these estimators are well documented by [13].

2.3 Approximate entropy

Entropy can be considered a measure of complexity of the considered system. The systems with high entropy can be characterized by no information flows and are thus random up to uncertainty and conversely, the series with low entropy can be seen as deterministic [22]. The efficient market can be then seen as the one with maximum entropy; and the lower the entropy, the less efficient the market is. For purposes of the Efficiency Index, we need an entropy measure which is bounded. Therefore, we utilize the approximate entropy introduced by [20].

For a random process consisting of independent random variables with identical uniform distributions, the approximate entropy converges to $-\log\left(r/\sqrt{3}\right)$ for all m [20]. For a completely deterministic process, the entropy goes to 0. Therefore, we can rescale the approximate entropy so that $0 \leq ApEn \leq 1$, where 0 characterizes a completely deterministic process and 1 a completely uncertain process characteristic for the efficient market[1]. In turn, it can be utilized in the Efficiency Index, a definition of which follows.

2.4 Capital market efficiency measure

[15] introduce the Efficiency Index (EI), defined as

$$EI = \sqrt{\sum_{i=1}^{n} \left(\frac{\widehat{M_i} - M_i^*}{R_i} \right)^2}, \tag{1}$$

where M_i is the ith measure of efficiency, $\widehat{M_i}$ is an estimate of the ith measure, M_i^* is an expected value of the ith measure for the efficient market and R_i is a range of the ith measure. In words, EI is simply a distance from the efficient market situation[2]. Here, we base the index on three measures of market efficiency – Hurst exponent H with an expected value of 0.5 for the efficient market ($M_H^* = 0.5$), fractal dimension D with an expected value of 1.5 ($M_D^* = 1.5$), and the approximate entropy with an expected value of 1 ($M_{AE}^* = 1$). The Hurst exponent is taken as an average of the GPH and the local Whittle estimates. In the same way, the fractal dimension is set as an average of the Hall-Wood and Genton estimates. For the approximate entropy, we utilize the estimate described in the corresponding section. The approximate entropy needs to be rescaled as it ranges between 0 and 1 with the efficient market of $ApEn = 1$. We thus have $R_{AE} = 2$ and $R_D = R_H = 1$.

[1] Note that the utilized estimator is a biased estimator of the true Approximate Entropy as well documented by [21]. Nevertheless, we opt to include it in the Efficiency Index as it controls for a different type of inefficiency than the fractal dimension and long-term memory do and it is the only entropy measure which is well bounded and thus scalable to a closed interval.

[2] Even though there might be various specifications of distance from the efficient market situation, we opt for this specific one as it is very straightforward.

3 Data description and results

We analyze daily prices of front futures, i.e., futures with the earliest delivery, of 25 commodities in the period between January 1, 2000 and July 22, 2013[3]. The dataset contains four energy types (Brent crude oil, WTI crude oil, heating oil, and natural gas), five metals (copper, gold, silver, palladium, and platinum), seven grains (corn, oats, rough rice, soybean meal, soybean oil, soybeans and wheat), five soft commodities (cocoa, coffee, cotton, orange juice, and sugar) and four other agricultural commodities (feeder cattle, lean hogs, live cattle, and lumber) futures from the Chicago Board of Trade (CBOT), Chicago Mercantile Exchange (CME), InterContinentalExchange (ICE), New York Mercantile Exchange (NYMEX), and its division Commodity Exchange (COMEX). We analyze logarithmic prices $S_{i,t} = \log P_{i,t}$, where $P_{i,t}$ is the price of futures i at time t, for the fractal dimension and logarithmic returns $r_{i,t} = S_{i,t} - S_{i,t-1}$ for the long-term memory and approximate entropy. The returns of all the analyzed futures are stationary according to ADF [6] and KPSS [17] tests.

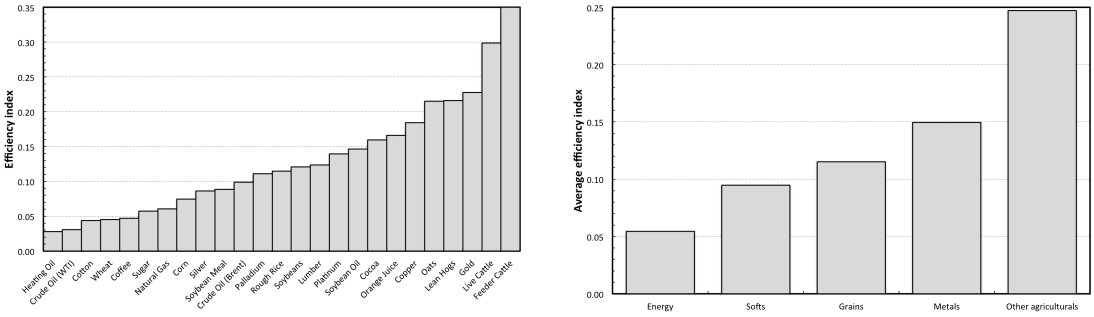


Figure 1 **Efficiency Index of commodity futures and groups of commodities.** In the left panel, the commodities are sorted from the most efficient to the least efficient. In the right panel, the commodities are clustered according to their type.

The estimated values of Hurst exponents, fractal dimensions and approximate entropies imply the following results. We observe that a fractal dimension below 1.5, which indicates local persistence, is characteristic for the majority of commodities. These series are thus locally trending. This is most evident for feeder cattle, lean hogs, and live cattle, i.e., mostly livestock futures. On the contrary, the energy commodities – namely both crude oils and natural gas – are close to a fractal dimension of 1.5 and, as such, they do not show any signs of local inefficiencies. For the long-term memory part, most of the futures are below 0.5 indicating anti-persistence which translates into a mean-reversion of prices, something that is not commonly observed for stocks, stock indices and exchange rates which are characterized by a unit-root behavior. The strongest anti-persistence is seen for cocoa, oats, and orange juice. Nonetheless, there is a portion of commodities which shows signs of persistence. These are copper, palladium, platinum, and sugar. Cotton and natural gas get the closest to the value of the efficient market. For the approximate entropy, several values are close to 1 for the efficient market[4] – lumber, sugar, and heating oil. The most complex, and thus the least efficient, series include feeder cattle and live cattle.

Putting these results together, we arrive at the Efficiency Index and efficiency ranking values which are graphically represented in Fig. 1. The most efficient of the commodities turns out to be heating oil, closely followed by WTI crude oil. Cotton, wheat, and coffee come after these with a similar level of efficiency. The ranking is then supplemented by other commodities, the efficiency of which increases quite steadily across the ranking. Feeder cattle is the least efficient commodity in this set, quite closely followed by live cattle. The livestock futures thus seem to be rather inefficient compared to the others. Connected to this finding, we also show an average

[3]The time series were obtained from http://www.quandl.com server on July 23, 2013.
[4]Several values even reach a value above 1 due to the finiteness of the sample.

Efficiency Index value for commodities according to their type. In the figure, we can see that the energy futures are the most efficient followed by soft commodities, grains, and metals. By far the least efficient group consists of the other agricultural commodities, i.e., feeder cattle, lean hogs, live cattle, and lumber. This is well in hand with the observations about very inefficient livestock futures.

4 Conclusions

We have analyzed the market efficiency of 25 commodity futures across various groups – metals, energies, soft commodities, grains, and other agricultural commodities. To do so, we have utilized the recently proposed Efficiency Index to find that the most efficient of all the analyzed commodities is heating oil, closely followed by WTI crude oil, cotton, wheat and coffee. On the other (least efficient) end of the ranking scale, we have detected live cattle and feeder cattle. The efficiency also seems to be characteristic for specific groups of commodities – energy commodities have been found the most efficient, followed by soft commodities, grains, and metals, whereas the other agricultural commodities (formed mainly of livestock) form the least efficient group.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. FP7-SSH-612955 (FinMaP) and the Czech Science Foundation project No. P402/12/G097 "DYME – Dynamic Models in Economics".

References

[1] Barunik, J., Kristoufek, L.: On Hurst exponent estimation under heavy-tailed distributions. *Physica A* **389**, 18 (2010), 3844–3855.

[2] Beran, J.: *Statistics for Long-Memory Processes*. Monographs on Statistics and Applied Probability, vol. 61, Chapman and Hall, New York, 1994.

[3] Cont, R.: Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **1**, 2 (2001), 223–236.

[4] Di Matteo, T.: Multi-scaling in finance. *Quantitatice Finance* **7**, 1 (2007), 21–36.

[5] Di Matteo, T., Aste, T., Dacorogna, M.: Scaling behaviors in differently developed markets. *Physica A* **324** (2003), 183–188.

[6] Dickey, D., Fuller, W.: Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74** (1979), 427–431.

[7] Elton, E., Gruber, M., Brown, S., Gotzmann, W.: *Modern Portfolio Theory and Investment Analysis*. John Wiley & Sons, New York, 2003.

[8] Fama, E.: The behavior of stock market prices. *Journal of Business* **38** (1965), 34–105.

[9] Fama, E.: Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance* **25** (1970), 383–417.

[10] Fama, E.: Efficient Capital Markets: II. *Journal of Finance* **46**, 5 (1991), 1575–1617.

[11] Geweke, J., Porter-Hudak, S.: The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**, 4 (1983), 221–238.

[12] Gneiting, T., Schlather, M.: Stochastic Models That Separate Fractal Dimension and the Hurst Effect. *SIAM Review* **46**, 2 (2004), 269–282.

[13] Gneiting, T., Sevcikova, H., Percival, D.: *Estimators of fractal dimension: Assessing the roughness of time series and spatial data.* Technical report, Department of Statistics, University of Washington, 2010.

[14] Kristoufek, L.: How are rescaled range analyses affected by different memory and distributional properties? A Monte Carlo study. *Physica A* **391** (2012), 4252–4260.

[15] Kristoufek, L., Vosvrda, M.: Measuring capital market efficiency: Global and local correlations structure. *Physica A* **392** (2013), 184–193.

[16] Kristoufek, L., Vosvrda, M.: Commodity futures and market efficiency. *Energy Economics* **42** (2014), 50–57.

[17] Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y.: Testing the null of stationarity against alternative of a unit root: How sure are we that the economic time series have a unit root? *Journal of Econometrics* **54** (1992), 159–178.

[18] Malkiel, B.: The efficient market hypothesis and its critics. *Journal of Economic Perspectives* **17**, 1 (2003), 59–82.

[19] Phillips, P. C. B., Shimotsu, K.: Local whittle estimation in nonstationary and unit root cases. *The Annals of Statistics* **32**, 2 (2004), 659–692.

[20] Pincus, S.: Approximate entropy as a measure of system complexity. In: *Proceedings of the National Academy of Sciences of the United States of America* **88** (1991), 2297–2301.

[21] Pincus, S.: Approximate entropy (ApEn) as a complexity measure. *Chaos* **5** (1995), 110–117.

[22] Pincus, S., Kalman, R.: Irregularity, volatility, risk, and financial market time series. In: *Proceedings of the National Academy of Sciences of the United States of America* **101**, 38 (2004), 13709–13714.

[23] Robinson, P.: Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* **23**, 5 (1995), 1630–1661.

[24] Samuelson, P.: Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review* **6** (1965), 41–49.

[25] Taqqu, M., Teverosky, W., Willinger, W.: Estimators for long-range dependence: an empirical study. *Fractals* **3**, 4 (1995), 785–798.

[26] Taqqu, M., Teverosky, W.: On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series. In: *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*, 1996.

[27] Teverosky, W., Taqqu, M., Willinger, W.: A critical look at lo's modified r/s statistic. *Journal of Statistical Planning and Inference* **80**, 1-2 (1999), 211–227.

On the design accuracy of Royall's predictor of domain total for longitudinal data

Małgorzata K. Krzciuk[1]

Abstract. In the paper the small area estimation problem is considered under noninformative sampling designs, in the context of predicting domain total for longitudinal data. The predictor proposed by Royall [5] is considered. It is the best linear unbiased predictor for any given sample. If the sampling design is noninformative it has the smallest model expected design mean squared error as well. Hence, it is known that it has good properties based on the model approach but also taking into account both design and model approaches as well. In the paper we study only its design-based properties to show the influence of the choice of the sampling design on the design accuracy of the predictor. We study simple random sampling without replacement, stratified sampling, sampling design proposed by Midzuno [4] and Lahiri [3] and two-stage sampling. For each sampling design we consider the case of the balanced panel sample. The Monte Carlo simulation study is based on the real longitudinal data on the number of new-registered companies and the registered unemployment rate in Polish poviats divided into 32 subpopulations.

Keywords: sampling design, Royall's predictor, longitudinal data, simulation.

JEL Classification: C42
AMS Classification: 97K80

1 Introduction

We consider the problem of prediction of the domain total for longitudinal data. Special case of the Linear Mixed Model is considered, similar to the model proposed by Battese, Harter and Fuller [1] but for longitudinal data and with profile specific instead of domain specific random effects. Design-based properties of the special case of the Royall's [5] best linear unbiased predictor are studied in the Monte Carlo simulation study.

2 Model-based and design-based approach in survey sampling

Two main approaches are considered in the survey sampling: model-based and design-based approach. In the model-based approach it assumed that the values of the variable of interest are realizations of random variables. The joint distribution of the random variables is denoted by ξ and it is assumed that it belongs to some class of distributions – the assumptions are called the superpopulation model (cf. Cassel et al. [2], p. 80). In the model approach the prediction accuracy is measured by taking into account different realizations of the superpopulation model (as in econometrics). In the design-based approach the values of the variable of interest are assumed to be fixed. The estimation accuracy is measured under the repeated sampling (taking into account the sampling design).

Sampling design of the sample s is a function $p(s)$, on the sample space Q, which – for all samples $s \in Q$ – meets the conditions $p(s) \geq 0$ and $\sum_{Q} p(s) = 1$. When we assume that S is the random variable with values $s \in Q$, then $P(S=s) = p(s)$, therefore the sampling design describe discrete distribution of S (cf. Cassel et al. [2], pp. 8-9, Wywiał [9], p. 17). Sampling scheme is the algorithm which enables to select the population elements to the sample according to the probabilities defined by sampling design (cf. Wywiał [9], p. 20).

Sampling design, denoted by $p(s)$, is called noninformative when function $p(.)$ does not depend on values of the response variable (y) (Cassel et al. [2], pp. 12-13). Let us take into account two sources of randomness in survey sampling: the distribution ξ – the joint distribution of the vector of random variables $\mathbf{Y}$ and the sampling design. If the sampling design is noninformative then each ξ-unbiased or p-unbiased predictor $\hat{\theta}$ is $p\xi$-

[1] University of Economics in Katowice, 1 Maja 50, Katowice, Poland, mkrzciuk@ue.katowice.pl.

unbiased. It is worth noting that the order of the operators E_p and E_ξ can be changed for noninformative sampling designs. Hence (cf. Cassel C. et al. [2], pp. 90-94),

$$E_p E_\xi\left(\hat{\theta}-\theta\right) = E_\xi E_p\left(\hat{\theta}-\theta\right) \tag{1}$$

and

$$\begin{aligned} E_p E_\xi\left(\hat{\theta}-\theta\right) &= E_p\left(0\right) = 0\,, \\ E_p E_\xi\left(\hat{\theta}-\theta\right) &= E_\xi E_p\left(\hat{\theta}-\theta\right) = E_\xi\left(0\right) = 0\,, \end{aligned} \tag{2}$$

respectively for ξ-unbiased or p-unbiased predictor $\hat{\theta}$. Additionally, for noninformative sampling designs the following implication holds:

$$\forall_s E_\xi\left(\hat{\theta}-\theta\right)^2 = \min \Rightarrow E_p E_\xi\left(\hat{\theta}-\theta\right)^2 = \min\,. \tag{3}$$

It means, that if the sampling design is noninformative then the predictor which minimizes the prediction mean square error for each given sample s minimizes the ξ-expected value of the p-mean squared error, too. Therefore, to find the predictor which minimizes ξ-expected value of the p-mean square error we can look for the predictor which minimizes the prediction mean square error for any given sample s (Żądło [10], pp. 29-30).

The general linear model (for the surveys in one period) is defined as follows (Valliant et al. [8], p. 26-27):

$$\begin{aligned} E_\xi\left(\mathbf{Y}\right) &= \mathbf{X}\boldsymbol{\beta} \\ D^2_\xi(\mathbf{Y}) &= \mathbf{V}\,, \end{aligned} \tag{4}$$

where $\mathbf{Y}$ is the vector of random variables, $\mathbf{X}$ is the matrix of auxiliary variables, of size $N\times p$ and $\boldsymbol{\beta}$ is vector of p unknown parameters. Positive definite covariance matrix is denoted by $\mathbf{V}$. Let the population elements are rearranged and first n rows of $\mathbf{X}$ are for units in the sample and first n elements of $\mathbf{Y}$ are those in the sample, then the matrixes $\mathbf{X}$ and $\mathbf{V}$ can be written as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{bmatrix} \quad \mathbf{V} = \begin{bmatrix} \mathbf{V}_{ss} & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_{rr} \end{bmatrix},$$

where $\mathbf{X}_s$ is of size $n\times p$ and $\mathbf{X}_r$ is of size $(N-n)\times p$ and submatrixes $\mathbf{V}_{ss}$, $\mathbf{V}_{rr}$, $\mathbf{V}_{rs} = \mathbf{V}_{sr}^{'}$ are of sizes $n\times n$, $(N-n)\times(N-n)$, $(N-n)\times n$, respectively.

The Best Linear Unbiased Predictor (BLUP) of $\theta = \boldsymbol{\gamma}^T\mathbf{Y}$ (where $\boldsymbol{\gamma}$ is the vector of constants), according to the general prediction theorem (Royall [5]) is the predictor that minimizes the error variance among linear, prediction unbiased estimators $\hat{\theta}$. For the model (4) this predictor is given by:

$$\hat{\theta}_{opt} = \boldsymbol{\gamma}_s'\mathbf{Y}_s + \boldsymbol{\gamma}_s'\left[\mathbf{X}_r\hat{\boldsymbol{\beta}} + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\left(\mathbf{Y}_s - \mathbf{X}\hat{\boldsymbol{\beta}}\right)\right], \tag{5}$$

where $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{Y}_s$, and $\boldsymbol{\gamma} = \left[\boldsymbol{\gamma}_s^T \;\; \boldsymbol{\gamma}_r^T\right]^T$, and $\boldsymbol{\gamma}_s$ is a subvector of $\boldsymbol{\gamma}$ for sampled elements.

3 Examples of sampling plans

In this section we present some sampling designs for surveys conducted in one time period: simple random sampling without replacement, Midzuno [4] and Lahiri [3] sampling design, stratified sampling and two-stage sampling.

The simple random sampling without replacement is the sampling design defined as follows (cf. Särndal, Swensson, Wretman [6], pp. 27-32):

$$p(s) = \binom{N}{n}^{-1}, \tag{6}$$

where N is the population size and n is the sample size. The first element, according to the sampling algorithm, is selected with probability $q_1 = \frac{1}{N}$ and the k-th element (where $k = 2, 3, ..., n$) with probability $q_k = \frac{1}{N - k + 1}$ (Wywiał [9], p. 21).

Let us consider the sampling design studied by Midzuno [4] and Lahiri [3]. It is given by:

$$p(s) = \frac{\bar{x}}{\binom{N}{n}\bar{X}}, \tag{7}$$

where $\bar{X}$ is the population mean of the auxiliary variable and $\bar{x}$ is the sample mean of the variable (Scott, Wu [7], p. 102). Based on the sampling scheme studied by Midzuno [4] and Lahiri [3] the first population element is selected with probability proportional to the value of the auxiliary variable:

$$p(k_1) = \frac{x_{k_1}}{\sum_{i=1}^{N} x_i} \tag{8}$$

for $k = 1, ..., N$. Other elements are selected as the simple random sample without replacement of size $n-1$.

The next sample design considered in the paper is the stratified sampling. It is assumed that population U is divided into H non-empty, separable subsets called strata (U_h, $h = 1, ..., H$ and $U = \bigcup_{h=1}^{H} U_h$). This sampling design can be written as follows (Wywiał [9], s. 101-102):

$$p(s) = \prod_{h=1}^{H} \binom{N_h}{n_h}^{-1}, \tag{9}$$

where N_h is the number of elements in h-th strata in the population and n_h is the number of elements sampled from the strata ($1 \le n_h \le N_h$). Hence, $N = \sum_{h=1}^{H} N_h$ and $s = \{s_1, s_2, ..., s_H\}$ where s_h, for $h = 1, 2, ..., H$, is a simple random sample without replacement selected from the strata h.

The last sampling design is the two-stage sampling. We assume the division of the population into non-empty and separable subsets. In this case we divide the population U into G groups denoted by U_p ($p = 1, ..., G$). The size of the p-th group and the sample size in the group are denoted by N_p and n_p, respectively. The sampling design is given by (Wywiał [9], pp. 115-120):

$$p(s) = \binom{G}{g}^{-1} \prod_{p=1}^{G} \binom{N_p}{n_p}^{-1}. \tag{10}$$

In the first stage we choose g from G groups. Each group is chosen to the sample with equal probability. Then, in the second stage, we choose simple random sample without replacement of the elements from the sampled groups. In the paper we study self-weighting two-stage design what means that the same fraction f of elements is selected at the second stage i.e. $n_p = f \cdot N_p$.

In the simulation study in the section 5 we will study balanced panel samples what means that in the first period a sample will drawn at random (using one of the sampling designs presented above) and then the same elements will be chosen to the sample in the next periods.

4 Predictor of domain total for linear mixed models

In this paper we study the following longitudinal nested error regression model (Żądło [11], pp. 91-92):

$$Y_{idj} = \mathbf{x_{idj}}\boldsymbol{\beta} + v_{id} + e_{idj}\,. \tag{11}$$

In the model, Y_{idj} is the random variable, $\mathbf{x_{idj}} = \left[x_{idj1} \quad x_{idj2} \quad ... \quad x_{idjp} \right]$ is the vector of p values of the auxiliary variables, where i is the number of the population element, d – the number of the domain and j – the number of the period ($i = 1, 2, ..., N \quad d = 1, 2, ..., D \quad j = 1, ..., t$). We consider, therefore, the case that the population U is divided into D subpopulations called domains (U_d). Number of elements in the d-th domain in the population is N_d. Additionally, v_{id} and e_{idj} are mutually independent, furthermore $v_{id} \sim (0, \sigma_v^2)$ and $e_{idj} \sim (0, \sigma_e^2)$.

The assumption (11) implies that:

$$Cov_\xi(Y_{idj}, Y_{i'j'd'}) = \begin{cases} 0 & \text{if} \quad i \neq i' \vee d \neq d' \\ \sigma_e^2 + \sigma_v^2 & \text{if} \quad i = i' \wedge j = j' \\ \sigma_v^2 & \text{if} \quad i = i' \wedge d = d' \wedge j \neq j' \end{cases}, \tag{12}$$

where the vector of unknown variance parameters has the form: $\boldsymbol{\delta} = \left[\sigma_e^2 \quad \sigma_v^2 \right]^T$. It should be noted that this model is similar to the model proposed in Battese, Harter and Fuller [1]. These authors, however, considered the model for only one period and with domain-specific random effects. In the paper the model is used for longitudinal data and takes into account profile-specific random effects.

The best linear unbiased predictor of the domain total based on the Royall's [5] theorem for the model (11) is given by (Żądło [11], pp. 96-99)

$$\hat{\theta}_{BLU} = \sum_{i \in s_{dt}} Y_{idt} + \sum_{i=1}^{N_{rdt}} \mathbf{x_{idt}}\hat{\boldsymbol{\beta}} + \sigma_v^2 \sum_{i=1}^{N_{rdt}} b_{id}^{-1} \sum_{j=1}^{m_{id}} \left(Y_{idj} - \mathbf{x_{idj}}\hat{\boldsymbol{\beta}} \right), \tag{13}$$

where $b_{id} = \sigma_e^2 + \sigma_v^2 m_{id}$ and m_{id} is the number of observations of the i-th element in the d-th domain, furthermore, N_{rdt} is the number of elements not selected into the sample in the d-th domain and t-th period. The vector $\hat{\boldsymbol{\beta}}$ in this case is given by (Żądło [11], pp. 96-98):

$$\hat{\boldsymbol{\beta}} = \left(\sum_{d=1}^{D} \sum_{i=1}^{n_d} b_{id}^{-1} \mathbf{X_{sid}^T} \mathbf{X_{sid}} \right)^{-1} \left(\sum_{d=1}^{D} \sum_{i=1}^{n_d} b_{id}^{-1} \mathbf{X_{sid}^T} \mathbf{Y_{sid}} \right) \tag{14}$$

Under the balanced panel sample the predictor (13) simplifies to:

$$\hat{\theta}_{BLU} = \sum_{i \in s_{dt}} Y_{idt} + \hat{\boldsymbol{\beta}} \sum_{i=1}^{N_{rdt}} \mathbf{x_{idt}}\,, \tag{15}$$

where:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{d=1}^{D} \sum_{i=1}^{n_d} \mathbf{X_{sid}^T} \mathbf{X_{sid}} \right)^{-1} \left(\sum_{d=1}^{D} \sum_{i=1}^{n_d} \mathbf{X_{sid}^T} \mathbf{Y_{sid}} \right) \tag{16}$$

and

$$\sum_{i=1}^{N_{rdt}} \mathbf{x_{idt}} = \left[\sum_{i=1}^{N_{rdt}} x_{idt1} \quad \sum_{i=1}^{N_{rdt}} x_{idt2} \cdots \quad \sum_{i=1}^{N_{rdt}} x_{idtp} \right]. \tag{17}$$

The known matrix $\mathbf{X_{sid}}$, of size $m_{id} \times p$, includes values of the auxiliary variables and $\mathbf{Y_{sid}}$ is vector of the values of the dependent variable. Elements of $\sum_{i=1}^{N_{rdt}} \mathbf{x_{idt}}$ are sums of auxiliary variables for domain elements not selected into the sample.

The model (12) and the predictor (14) can be generalized assuming spatial correlation of random effects and temporal correlation of random components (see Żądło [12]).

5 Simulation study

In this section results of the simulation studies are presented. In the study we use data from Local Data Bank (Polish Central Statistical Office). We consider a longitudinal data for Polish regions called poviats in years 2010-2012. In the analyzed model the dependent variable is the number of new-registered entities of the national economy included in the REGON register. The auxiliary variable is the registered unemployment rate in years 2009-2011. In the model we take into account the intercept. The division of the population of poviats (of size $N = 379$ in each out of 3 periods) into $D = 32$ subpopulations is made according to the larger regions called voivodships and the dominating place of residence of citizens ("city" or "village").

In each design-based simulation study the balanced panel sample is considered. We sample a sample of size $n = 36$ in the first period and the same elements are observed in the second and in the third period. In each simulation study 5000 samples are drawn at random using different sampling schemes. For two-stage sampling we assume $g = 4$ and $f = 0.25$.

Some of the functions used in the simulation study are available in the *sampling* package: *srswor* (for simple random sampling without replacement in one period), *UPmidzuno* (for Midzuno and Lahiri sampling plan in one period) and *strata* (for stratified random sampling in one period). Function for two-stage sampling is prepared by the author as well as the balanced panel sample generalizations of all of the functions and, moerover, the function which allows to compute the value of the predictor.

In the simulation study the design-based mean squared error of the predictor for each out of $D = 32$ domains is computed as:

$$MSE_p(\hat{\theta}^{(d)}) = \frac{1}{5000} \sum_{l=1}^{5000} \left(\hat{\theta}_l^{(d)} - \theta^{(d)} \right)^2, \tag{18}$$

where $\hat{\theta}_l^{(d)}$ is the predictor of domain total for the *d*-th domain and the *l*-th iteration of the simulation study (given by (15)), $\theta^{(d)}$ is the fixed domain total.

In the Figure 1 we present the values of ratios of mean squared errors for sampling designs considered in the paper for balanced panel sample. Let us denote the value of the design mean squared error for balanced panel sample drawn using simple random sampling withour replacement by MSE(SRSWOR). Ratios of MSEs for Midzuno-Lahiri sampling design (on figure *MSE(ML)*) and simple random sampling without replacement are close to 1 (minimum value of the ratio is 0.984 and maximum is 1.133). For two-stage sampling (*MSE*(*2ST*)), the ratios of MSEs are from 1.040 to 2.974. In the case of stratification the median of the considered ratio (on figure labeled *MSE(strat)*) is close to 1 but minimum value is equal 0.657 and maximum – 1.276.

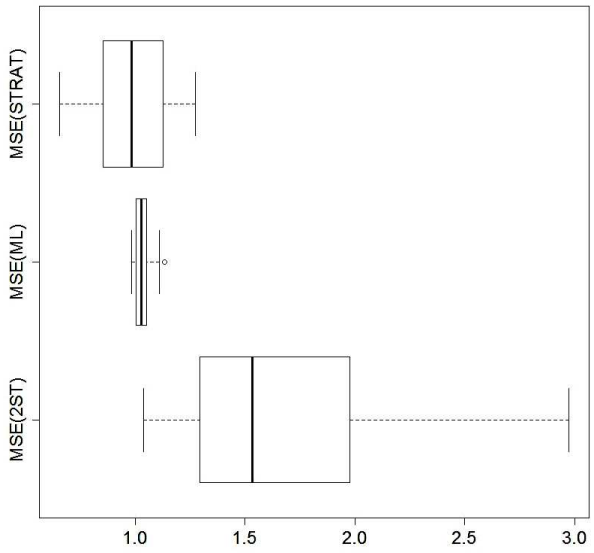


Figure 1 Values of ratios MSE(.)/MSE(SRSWOR) for D=32 domains

Hence, the choice of sampling design may have significant influence on the design-based properties of the predictor. The stratified random sampling gives in the simulation study even more than 30% (ratio of MSEs 0.657) gain in the design-based accuracy while choosing two stage sampling may cause even almost 3 times larger
p-mean squared errors.

6 Summary

In the paper we consider the problem of the design accuracy of Royall's predictor of domain total for longitudinal data. We study, in the Monte Carlo simulation, the influence of the choice of the sampling design on the design-based properties of the predictor. In the analyses we use real longitudinal data from Local Data Bank (Central Statistical Office). We analyze special case of the Linear Mixed Model – the modification of the model proposed by Battese, Harter and Fuller [1]. Obtained results of the simulation study suggest that the choice of the sampling design may have significant influence on the design-based properties of the considered predictor.

References

[1] Battese, G.E., Harter, R.M., Fuller, W.A.: An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* **83** (1988), 28-36.
[2] Cassel, C. et al.: *Foundations of Inference in Survey Sampling*. John Wiley & Sons, New Jersey, New York, 1977.
[3] Lahiri, G. W.: A method for Sample Selection Providing Unbiased Ratio Estimator. *Bulletin of the International Statistical Institute* **33** (1951), 133-144.
[4] Midzuno, H.: An outline of the theory of sampling systems. *Annals of the Institute of Statistical Mathematics* **1** (1950), 149-156.
[5] Royall, R.M.: The linear least squares prediction approach to two- stage sampling. *Journal of the American Statistical Association* **71** (1976), 657-664.
[6] Särndal, C.E., Swensson, B., Wretman, J.H.: *Model Assisted Survey Sampling.* Springer Verlag, New York, 1992.
[7] Scott, A.J., Wu, C.: On the asymptotic distribution of ratio and regression estimators. *Journal of the American Statistical Association* **76** (1981), 98-102.
[8] Valliant, R. et al.: *Finite Population Sampling and Inference*. John Wiley & Sons, New York, 2000.
[9] Wywiał, J. L.: *Wprowadzenie do metody reprezentacyjnej*. Wyd. Akademii Ekonomicznej w Katowicach, Katowice, 2010.
[10] Żądło, T.: *Elementy statystyki małych obszarów z programem R*. Wyd. Akademii Ekonomicznej w Katowicach, Katowice, 2008.
[11] Żądło, T.: On some problems of prediction of domain total in longitudinal surveys when auxiliary information is available. *Studia Ekonomiczne* **133** (2013), 86-106.
[12] Żądło, T.: On the prediction of the subpopulation total based on spatially correlated longitudinal data. *Mathematical Population Studies* **21** (2014), 30-44.

Mathematical models of the work contour in project management

Petr Kučera [1], Jan Bartoška [2], Tomáš Šubrt [3]

Abstract. The paper deals with the "Student Syndrome" phenomenon and its expression using different mathematical models. Every existing task in any project is, to a higher or lower extent, determined by the effect of the human agent, which commonly takes the Student's syndrome form. The inefficiency of a number of projects in practice is largely caused by an unsuccessful realization of partial tasks. The effect of the human agent is in this respect fundamental. The human agent, as an allocated resource in the task, is liable to a number of non-specified impacts and stimuli, and as such s/he is rather versatile in his/her behaviour. The versatility of the human agent in projects can be described by the "Student Syndrome" phenomenon. If the "Student Syndrome" phenomenon could be completely expressed by means of a mathematical model, it would be possible to improve the prediction of its impact on project tasks, and therefore to exploit the resources allocated in the project more efficiently and prevent project crises. We can not expect that the "Student Syndrome" phenomenon can be expressed in a trivial way and using only one mathematical function. This paper proposes several functions for this representation. The proposed functions are compared in terms of their course and the possibility of work schedule resources parameterization. We can regard this original viewpoint as suggestive for the area of human resources management in projects. The article brings the quantification of qualitative features of the human agent in project management.

Keywords. Project management, Student Syndrome, Parkinson's law, mathematical model, work effort, resource stimulation.

JEL Classification: C61
AMS Classification: 90B99

1 Introduction

Using a project time analysis, it is possible to apply the two most well-known methods, i.e. the Critical Path Method and the Critical Chain Method, whose comparison and the way of use is dealt with by Lechler et al [6], while Rand [9] writes about their preferences as well as their critics. Greater attention is given to the Critical Chain Method, which derives from the basic presuppositions of the Theory of Constraints. These presuppositions and starting points are also introduced by Leach [5], one of them being the phenomenon referred to as the "Student Syndrome" or "procrastination", which is the topic of this article. The research of procrastination, i.e. meeting deadlines at the last minute, and its elimination in university or higher education courses is dealt with by Tadepalli et al [11]. The "Student Syndrome" can occur during any human activity, especially when fulfilling tasks during the work on project. In their article, Zika-Viktorsson and Ingelgård [12] concentrate on the research of psychological aspects in project work environment.

The versatility of the human agent in projects can be described also by the first "Parkinson's law" [8]. It is natural for people to distribute work effort irregularly to the whole time sequence which was determined by the deadline of the task termination. The questions of "Parkinson's first law" in project management are further dealt with in e.g. [10].

[1] Czech University of Life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Praha 6, kucera@pef.czu.cz.
[2] Czech University of Life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Praha 6, bartoska@pef.czu.cz.
[3] Czech University of Life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Praha 6, subrt@pef.czu.cz.

Work effort of an allocated resource has very often been researched in projects from the area of informational technologies and software development, as these projects contain a high level of indefiniteness, and even common and routine tasks are unique. At the same time, it concerns the area where it is possible to find a great number of approaches to estimate how laborious the project will be or how long the tasks will take, and also case studies. The proposal for mathematical apparatus for planning the course of tasks within a case study is dealt with for instance in [7], or Barry et al. [1]. The authors Özdamar and Alanya [7] propose a particular pseudo-heuristic approach to estimate the tasks course where the indefiniteness in the project is expressed by fuzzy sets. Barry et al. [1] concentrate on the existence and expression of the relation between project duration and total effort and in their theoretical starting points they point out the dynamics of the relation between the effort and project duration when a self-strengthening loop can be expected. The others who research the project complexity and work effort are for instance Clift and Vandenbosh [3], who point out a connection between the length of life cycle and project management structure where a key factor is again a human agent.

The resource allocation should always proceed from the resource effort which is not constant during task performance. The work effort will be different for various resources and tasks. The aim of the paper is to give a mathematical expression of the human factor impact in the form of the "Student Syndrome" for different types of work contours of tasks in projects

2 Materials and methods

2.1 Student Syndrom and Parkinson's law

Out of his/her natural character, man will always be inclined to relaxation during work effort, and to saving his/her vital energy. Unless a human being is exposed to a certain amount of stress from the loss of profit for a performed task, s/he is not motivated to suppress this for his/her natural behaviour. For a worker it is convenient to relax provided there is no threat of the loss of his/her future reward. In his book, "The Laws of Professor Parkinson" [8], Parkinson, in a hilarious way, selected social features having a negative impact are described. In his essays and observations the author especially mentions that "Work increases chronologically with the time that we are able to devote to it". It can be best demonstrated by the following: The one who has a whole day's worth of time is also the busiest one.

While assessing a work task, if the deadline of termination is assessed and its resource is a human agent, the resource uses his/her work effort during the realization of the work process irregularly and with a variable intensity. This delay in the realization process which human resources participate in, leads to stress, or to tension on the resource, or to the tension or stress of the resource him/herself. The human agent in the allocated resource evokes the increase in work effort as long as the tension develops and starts to grow. The "Student Syndrome" and its progress are dealt with in [4]. The concept is common in both fiction, as well as specialist literature, and it is also introduced in [5]. The concept developed on the basis of the observation of student's behaviour during their work on given tasks. We can find the equivalent of the phenomenon in psychology under the term procrastination.

2.2 Scheduling the work of resources

Resource conflicts can be resolved by many means – from overtime work allocation, postponing the beginning of work to resource substitution. For our work, we will focus on the method connected with the change in work contour. The contour of resource work for a task is given by at least three influence factors, which are as follows:

- Attitude to succumbing to the Student Syndrome [2];
- Real need for carried out work depending on the nature of a task with the resource working according to standard work contour, of the type *flat*, *back loaded*, *front loaded*, *early peak*, *late peak*, *double peak*, *bell* and *turtle*;
- Need to adjust one's work speed to key resources of a task because it can be expected that during the realization of various types of tasks various resources will have different work effort. By combining these factors we can expect that, in particular, complex tasks will have a different contour.

The real work contour of the task is the combination of basic contour and inclination to succumb to the Student Syndrome or Parkinson's Law. When solving resource conflicts, in contrast to prevailing experience, we propose the combination of basic contour with real resource effort which includes the impact of the human agent.

2.3 Mathematical model of the "Student Syndrome"

The work effort during the "Student Syndrome" is falling in the first half of the duration and is going up in the second half. Bartoška and Šubrt [2] propose a mathematical model for the "Student Syndrome" as the multiplication of a growing work trend (exponential function) and the effect of the resource allocation (linear formula with parameter *z*):

$$p = \left| \sin\left[8\pi \left(t - 0{,}5\right)^3 \right] \left(t + \frac{z-2}{2} \right) e^t \right| \tag{1}$$

where a dependent variable *p* is the estimate of real resource allocation in %, an independent variable *t* is the task duration in %, and *z* is the estimate of expected maximum resource allocation in % during task realization.

The model (see Formula 1) enables the modelling of the resource allocation for work effort during a task. By changing the parameter *z* in the range <0;2>, we can obtain the estimate of the variability of the resource work effort in time, i.e. [2]:

- Replacing 0% for the parameter *z* we expect inconsiderable or no resource allocation;
- Replacing 100% for the parameter *z* we expect an average or relevant resource allocation;
- Replacing 200% for the parameter *z* we expect extreme allocation or resource over allocation.

3 Results and discussion

3.1 Improvement of the mathematical expression of the "Student Syndrome"

First, we will concentrate only on the function of sinus in (1), i.e. the mathematical description of the proper expression of the "Student Syndrome". Let this function be denoted by p_1. The remaining part of (1) explaining the resource allocation and the growing work trend will be dealt in the following chapters.

The rate of resources utilization in the allocation of resources during the implementation of project tasks can not exceed 100%. We can not expect resources utilization outside of their power. Therefore, to express the "Student Syndrome", it is necessary to find such a function p_1 so that

$$\int_0^1 p_1(t)\,dt = 1 \tag{2}$$

For computation of the integral in (2), polynomials are more convenient than goniometric functions. Function p_1 should have three minima $p_1(t) = 0$ in $t = 0$, $t = 0.5$, and $t = 1$; and two maxima: former one close to the begin and latter one close to the end of the task realization. The description of the "Student Syndrome" which requires these conditions can be realized using the 4th degree polynomial:

$$p_1 = -120t^4 + 240t^3 - 150t^2 + 30t \tag{3}$$

3.2 Mathematical expression of standard work contours

In this chapter, functions denoted by p_2 expressing the resource allocation according to standard work contours are proposed. Similarly, for all these functions p_2:

$$\int_0^1 p_2(t)\,dt = 1 \tag{4}$$

The simplest situation is in cases back loaded and front loaded where linear functions (5) a (6), respectively, suffices:

$$p_2 = 2t \tag{5}$$

$$p_2 = 2 - 2t \tag{6}$$

For *double peak*, *bell*, and *turtle*, we propose both polynomial expressions (7), (8), and (9), respectively:

$$p_2 = -\frac{960}{13}t^4 + \frac{1920}{13}t^3 - \frac{1230}{13}t^2 + \frac{270}{13}t \tag{7}$$

$$p_2 = 30t^4 - 60t^3 + 30t^2 \tag{8}$$

$$p_2 = -6t^2 + 6t \tag{9}$$

3.3 Mathematical model combining the "Student Syndrome" and standard work contours

The expression of the "Student Syndrome" during the realization of a task can be variously strong. Therefore, we introduce the rate r of the "Student Syndrome" which acquires values between 0 and 1. The case of $r = 0$ will represent a situation when the "Student Syndrome" does not occur at all and the resource keep the work contour exactly. On the other hand, the case of $r = 1$ means that the "Student Syndrome" manifests in its all strength and the resource absolutely ignore the work contour. Of course, both the previous cases are hypothetical. As a result, we can model the resource work effort p during a real task realization in the following way:

$$p = rp_1 + (1-r)p_2 \tag{10}$$

From (2) and (4) the following holds again:

$$\int_0^1 p(t)\,dt = 1 \tag{11}$$

The following figures show instances of work effort during task realization for particular work contours and "Student Syndrome" rate.

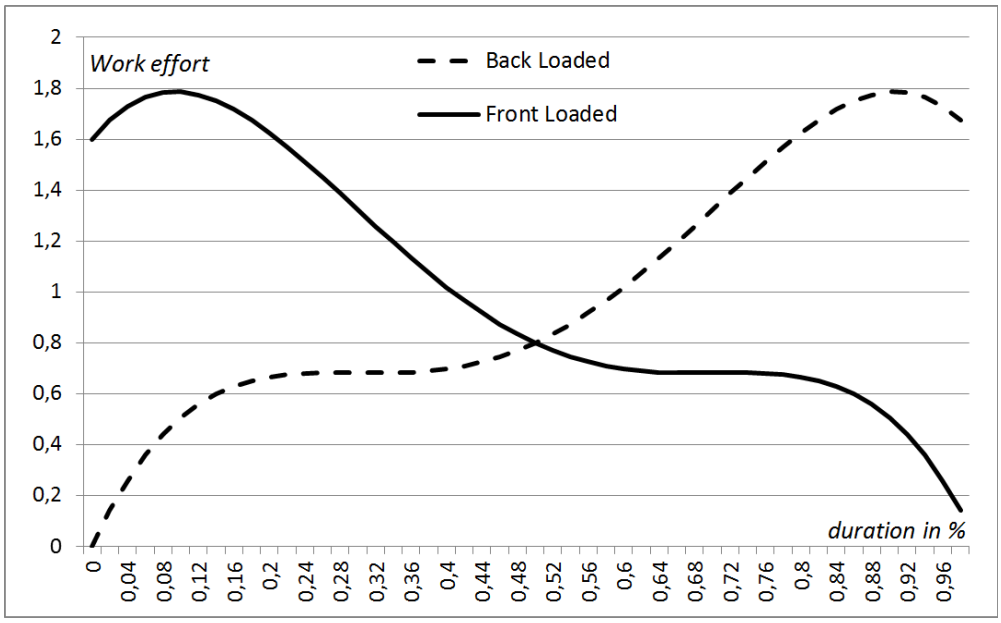


Figure 1 Back Loaded (5) and Front Loaded (6) with Student Syndrome ((10) with r = 0.2; $p = -24t^4+48t^3-30t^2+7.6t$ and $p = -24t^4+48t^3-30t^2+4.4t+1.6$, respectively)

If the human factor impact in the form of the "Student Syndrome" (12) is included into the work contour, the resource work process significantly changes. The impact on project tasks may be substantial.

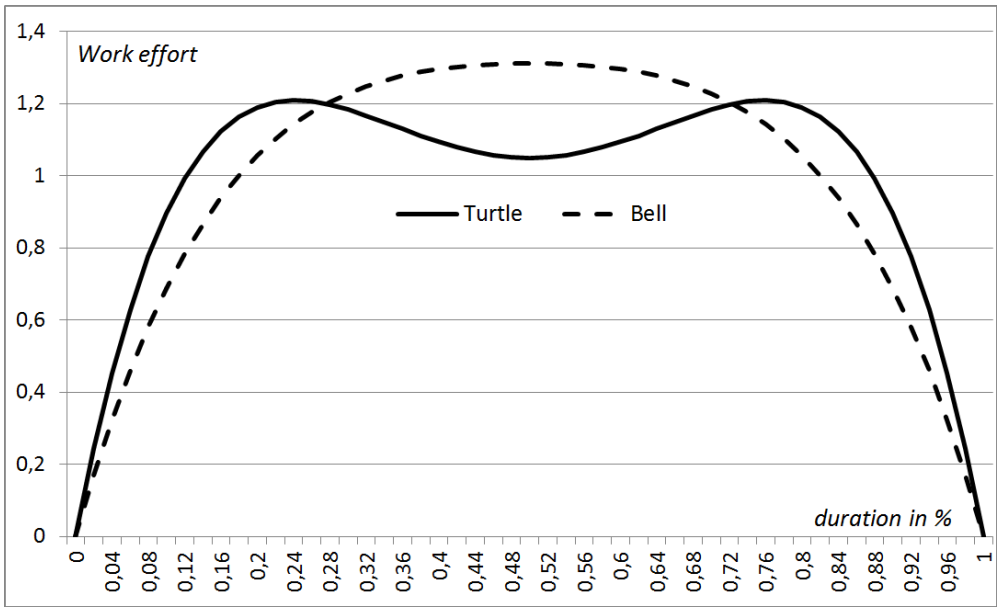


Figure 2 Bell (8) and Turtle (9) with Student Syndrome ((10) with r = 0.3; $p = -15t^4+30t^3-24t^2+9t$ and $p = -36t^4+72t^3-49.2t^2+13.2t$, respectively)

The mathematical models of work contours with the human resource impact can be implemented in planning the resource work in case of tasks of arbitrary project. Thereafter, different runs and shapes of the modified work contours can be composed and thus resource conflicts in projects can be eliminated.

4 Conclusion

The paper deals with the human resource impact in the project management. The particular contribution insists in proposals of mathematical models for the "Student Syndrome". These models are extended by selected resource work contours. Including the "Student Syndrome" impact, the work contour shape significantly changes. The proposal of the "Student Syndrome" impact enclosure into the work contour is based on the weight sum of two functions: the function of the "Student Syndrome" and the function of the work contour; both the functions are proposed here. The expression of the "Student Syndrome", i.e. the human resource impact, for a particular task and a particular resource in a project may be important in ordinary practice with the result of project costs savings or avoiding resource conflicts in project planning.

References

[1] Barry, E., Mukhopadhyay, T., Slaughter, A. S.: Software project duration and effort: An empirical study. *Information Technology and Management* **3** (2002), 113–136.

[2] Bartoška, J., Šubrt, T.: The effect of human agent in project management. *Central European Journal of Operations Research* **3** (2011), 369-382.

[3] Clift, T., Vandenbosh, M.: Project complexity and efforts to reduce product development cycle time. *Journal of Business Research* **45** (1999), 187–198.

[4] Cook, S.C.: *Applying Critical Chain to Improve the Management of Uncertainty in Projects* [M.B.A. thesis]. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Massachusetts, USA, 1998, 77 p.

[5] Leach, L. P. Critical Chain Project Management Improves Project Performance. *Project Management Journal* **30** (2) (1999), 39-51.

[6] Lechler, T. G., Ronen, B., Stohr, E. A. Critical Chain: A New Project Management Paradigm or Old Wine in New Bottles? *Engineering Management Journal* **17** (4) (2005), 45-58.

[7] Özdamar, L., Alanya, E.: Uncertainty modelling in software development projects (with case study). *Annals of Operations Research* **10** (2001), 157–178.

[8] Parkinson, C.N.: *Parkinson's Law and Other Selected Writings On Management*. 1st ed., Singapure, Federal Publications (S) Pte Ltd 1991.

[9] Rand, G. K. Critical chain: the theory of constraints applied to project management. *International Journal of Project Management* **18** (2000), 173-177.

[10] Gutierrez, G.J., and Kouvelis, P. Parkinson's law and its implications for project management. *Management Science* **37** (8) (1991), 990-1001.
[11] Tadepalli, S., Booth, C., Pryor, M. Evaluating academic procrastination in a personalized system of instruction based curriculum. In: *Proceedings of the ASEE Annual Conference and Exposition*, 2009.
[12] Zika-Viktorsson, A., Ingelgård, A. Reflecting activities in product developing teams: conditions for improved project management processes. *Research in Engineering Design* **17** (2006), 103–111.

Analysis of the Dependence of the Biggest Oil Companies' Revenues on the World Oil and Gas Price by using Panel Data Tools

Elena Kuchina [1]

Abstract. This paper is focused on the analyzing the dependence of the biggest oil companies' revenues on the World oil and gas price. The present study uses recently developed panel methodology to reveal the character of the relationships between considered variables. The paper examines if the difference in some control factors such as the headquarters' or parent company's location, the volume of the oil production causes the different responses of the revenues to the same changes in the oil and gas price. Panel data of 15 biggest oil companies from the 1Q2006 to the 3Q2013 was used in this study. The present paper takes into account the heterogeneity of panel with cross-section dependence, which is confirmed by the Pesaran (2004) CD test for cross-section dependence in panel data. Considering it, the second generations of the unit root test Pesaran (2007) as well as the second generation of the cointegration test Westerlund (2007) have been applied. The last part of this paper is ended by implementing panel estimators to determine the role of the control factors in the response of companies' revenues to the changes in the oil and gas price.

Keywords: panel data, cross-section dependence, unit root test, cointegration test.

JEL Classification: C33, G10
AMS Classification: 62-07

1 Introduction

The World oil and gas price is continually changing as a result of the influence of many different factors, which then have an impact on the demand and supply side. In some cases there can be recorded a so called oil shocks, which show in a huge fluctuation of the oil price. During the considered period from the 1st quarter of 2006 to the 3rd quarter of 2013 the largest increase in the oil price was recorded in July 2008, when the price of Brent crude oil reached the level of 132, 72 $/barel [7]. Then in December in the same year there was a very sharp drop in the oil price to 39, 95 $/barrel [7] due to the recession and as a result the falling oil demand [11].

It is obvious that such changes and shocks have a significant impact on many different industries in the business world, however, the most sensitive reactions can be expected in case of the oil companies, which revenues are straightly dependent on the world oil and gas price. Is the response of the biggest oil companies' revenues to the changes in price's fluctuations the same for all companies? The present study focused on the analysis if the differences in some control factors such as the headquarters' or parent company's location, or the volume of the oil production can cause the different companies' reaction in terms of their revenues to the same changes in the world oil and gas price.

The paper is structured into three parts. Firstly, the data set used in this research is described. Secondly, the methodology, determining the continuity of the steps in the empirical part, is introduced. The final chapter is focused on the empirical analyzing and modelling.

2 Data Set

This section discusses the input data used in the analysis. For the research the quarter data from 1st quarter of 2006 to the 3rd quarter of 2013 were used. In this analysis the data about quarter revenues of the world biggest oil companies with the headquarters' or parent's company location in Russia (LUKOIL, Rosneft, Gazprom),

[1] University of Economics, Prague, Econometrics dep., W Churchill Sq. 4, Czech Republic,
e-mail: xkuce10@vse.cz

Europe (BP, Royal Dutch Shell, Total, Eni, Statoil, OMV Group, PKN Orlen), in USA (ExxonMobile, Chevron Corporation, ConocoPhillips), in Mexico (Pemex) and in Brazil (Petrobras) were used. Most of these companies are on the list of the world's 25 biggest oil companies made by Forbes in 2012. The information about revenues of all companies was converted to one currency (USD). The data were obtained from the quarter financial statements of the mentioned companies, as well as, from Wikinvest [10].The data set considers other two variables: the world oil and gas price. The quarter data about oil price were obtained based on the monthly data [7]. In case of gas its world price was modeled by two types of the gas price: price of U.S. natural gas exports [8] and Natural Gas spot price at the Henry Hub terminal in Louisiana [2]. The quarter data was similarly gained based on the monthly data.

3 Methodology

There are two groups of the unit root test: "first generation unit root test" and "second generation unit root test". The main difference between these two groups consists in the cross section independence assumption allowing to use the first generation unit root test. The second generation unit root test takes into account the heterogeneity of panel with cross section dependence, which can be verified by Pesaran (2004) CD test [5]:

$$CD = \sqrt{\frac{2}{N(N-1)}}\left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{ij}}\, \hat{\rho}_{ij} \right),$$

where N - number of companies (15), T_{ij} – number of observations for which the correlation coefficient was computed, $\hat{\rho}_{ij}$ - the sample estimate par-wise correlation coefficients between the company i and j. This equation is appropriate for both balanced and unbalanced panels. The null hypothesis is cross section independence. The test is robust to nonstationarity, parameter heterogeneity or structural breaks. The Stata xtcd code used in this research allows to provide pre-estimation analysis of cross section dependence in the data, by its application to the variable series [1].

Following Pesaran (2007) [6] the second generation unit root test should be applied in case of cross- section dependence. In case of $AR(p)$ error specification the individual CADF statistics can be obtained from the following p^{th} order cross-section/time series augmented regression [6]:

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + c_i \bar{y}_{t-1} + \sum_{j=0}^{p} d_{ij} \Delta \bar{y}_{t-j} + \sum_{j=1}^{p} \delta_{ij} \Delta y_{i,t-j} + e_{it},$$

where $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$. $\bar{y}_{t-1}$ is a mean value of lagged levels.After calculating the individual *CADF* statistics, the *CIPS** statistics can be obtained by calculating their simple average. The null hypothesis assumes that all variables are non-stationary. The alternative hypothesis is a stationary process for at least one time series [6].

After having confirmed the nonstationarity of the considered variables the log-term dynamics can be tested by Westerlund (2007) [9] error-correction-based cointegration tests, which allow a large degree of heterogeneity and dependence within and across the cross sectional units. These tests assume the following data generated process [4]:

$$\Delta y_{it} = \delta'_i d_t + \alpha_i (y_{i,t-1} - \beta'_i x_{i,t-1}) + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{i,t-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta x_{i,t-j} + e_{it},$$

where $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$. For the deterministic component contained in d_t there are three cases: the process does not have deterministic terms, so $d_t = 0$; the generating process has a constant, so $d_t = 1$; the last case $d_t = (1,t)'$ assumes that the process contains both a constant and a trend. α_i determines the speed at which the system corrects back to the equilibrium relationship after sudden shock. The dependence across i is handled by means of bootstrap methods. Four cointegrated tests developed by Westerlund (2007) [9] test the null hypothesis of no cointegration. The alternative hypothesis of the two tests is that the panel is cointegrated as a whole, for the other two tests the alternative hypothesis is that at least one unit is cointegrated [4].

All mentioned analysis was applied on the logarithms of the considered variables. All computations were performed in software Stata/SE 11.2.

4 Empirical Modeling

According to the methodology mentioned above, the Pesaran (2004) CD test for cross-section dependence in panel data [2] was performed. The cross-dependence across units in case of variables related to the oil and gas price is obvious, as for all companies the world oil and gas price is considered the same. The following table (table1) contains the values of the CD test statistic as well as its *p*-value.

Variables	**CD Test Statistic**	***p*-value**
log Revenues	37,58	0,00

Table 1 Pesaran (2004) CD test for cross section dependence

Based on the *p*-value and CD Test Statistic in the table 1, the null hypothesis of cross-section independence across unit in panel data can be rejected on the 1% level of significance.

As the cross-section independence assumption is not met, the second generation unit root test should be applied. The table 2 and 3 report the results of the Pesaran *CADF* unit root test[3]. In the table 2 there are results of the constant specification only, while the table 3 represents the constant with trend specification. In order to take into account the possibility of the serial correlation in quarter data in both cases the p^{th} order augmented *CADF* regression was applied, where $p = 1,\ldots,4$.

	***p*=1**	***p*=2**	***p*=3**	***p*=4**
Variables	***t*-bar test (*p*-value)**	***t*-bar test (*p*-value)**	***t*-bar test (*p*-value)**	***t*-bar test (*p*-value)**
log Revenues	-2,477(0,002)	-1,844(0,380)	-1,632(0,715)	-1,529(0,840)
log Price Oil	2,610(1,000)	2,610(1,000)	2,610(1,000)	2,610(1,000)
log Gas Oil	2,610(1,000)	2,610(1,000)	2,610(1,000)	2,610(1,000)

Table 2 Pesaran *CADF* unit root test (constant)

	***p*=1**	***p*=2**	***p*=3**	***p*=4**
Variables	***t*-bar test (*p*-value)**	***t*-bar test (*p*-value)**	***t*-bar test (*p*-value)**	***t*-bar test (*p*-value)**
log Revenues	-3,092(0,000)	-2,157(0,789)	-1,704(0,997)	-1,563(1,000)
log Price Oil	1,700(1,000)	1,700(1,000)	1,700(1,000)	1,700(1,000)
log Gas Oil	1,700(1,000)	1,700(1,000)	1,700(1,000)	1,700(1,000)

Table 3 Pesaran *CADF* unit root test (constant and trend)

In both cases the null hypothesis of unit root existence cannot be rejected for all variables for the $p = 2, 3, 4$. For the $p = 1$ the null hypothesis can be rejected for the variable *log Revenues*. For the lags more than one all variables are non-stationary, but their first differences for the $p = 1,2$ are stationary, i.e. variables are I (1) on the 10% level of significance.

Cointegration analysis

Due to the evidence of the nonstationarity of the considered variables the cointegration analysis can be applied. Results of the panel cointegration test proposed by Westerlund (2007)[4] between the oil companies' revenues and oil price are reported on the table 4. The order of *lag* was chosen equal to one (in order to avoid overparametrization), the specification of the cointegration relationship was assumed the constant only as well as a constant and trend. In cases where the trend was statistically significant it was put into cointegration equation. The Bartlett kernel window width set was chosen equal to 3 based on the $4(T/100)^{2/9}$ with $T = 31$ [4]. Due to the

[2] In Stata *xtcd* code was used, created by Markus Eberhardt, University of Oxford Manor Road, Oxford OX1 3UQ.

[3] In Stata *pescadf* code was used, created by Piotr Lewandowski, Warsaw School of Economics, Institute for Structural Research, Poland.

[4] In Stata *xtwest* code was used, created by Damiaan Persyn, Katholieke Universiteit Leuven, Leuven, Belgium and Joakim Westerlund, Lund University Lund, Sweden.

presence of the cross section dependence across the units the robust critical values should be obtained through bootstrapping with 300 replications.

Statistics	*p*-value	Robust *p*-value	α_i
	Russia(trend)		
$G\tau$	0,006	0,020	
G_α	0,004	0,027	-0,8329
$P\tau$	0,000	0,017	
P_α	0,000	0,027	
	USA(trend)		
$G\tau$	0,001	0,027	
G_α	0,000	0,030	-0,7630
$P\tau$	0,690	0,643	
P_α	0,780	0,740	
	Europe(without trend)		
$G\tau$	0,000	0,000	
$G\alpha$	0,000	0,000	-0,6945
$P\tau$	0,000	0,000	
$P\alpha$	0,000	0,000	
	Mexico and Brazil[5]*(without trend)*		
$G\tau$	0,094	0,123	
$G\alpha$	0,077	0,103	-
$P\tau$	0,072	0,163	
$P\alpha$	0,018	0,133	

Table 4 Westerlund Error-correction-based panel cointegration test between revenues and oil price.

In case of test statistics $G\tau$ and $G\alpha$ rejection of H_0 indicates existence of cointegration of at least one of the cross-sectional units, while in case of test statistics $P\tau$ and $P\alpha$ rejection of H_0 denotes the evidence of the cointegration for the panel as a whole [4].

Based on the results in the table 4 the cointegration relationships between the logarithms of the companies' revenues and the oil price exist for all companies with the headquarters' or parent company's location in Russia and in Europe on the 5% level of significance. In case of the 2nd category, there is a conintegration relationship for at least one company with headquarter' location in USA. For companies with headquarters in Mexico and Brazil the cointegration relationship was not proven on the 5% level of significance. In case of location in Russia the system corrects back to the equilibrium after a sudden shock at speed equal to 0, 83. In case of location in USA this speed is a little lower but still remains relatively high, and is equal to 0,76, and in case of Europe this speed is equal to 0,69.

Next control variable examined in this paper is the volume of the oil production, which is measured in barrels of oil and natural gas equivalents. There were defined four categories[6], first category obtained 4 companies with the production volume less than 2 MBOE/D, in the second category there are 7 companies with the volume of production between 2 and 4 MBOE/D, the third category covers 3 companies with the volume less than 6 but more than 4 MBOE/D and in the last category there is only one company with the whole volume of the oil and natural gas equivalents production equal to 8, 1 MBOE/D. For each category the cointegration test Westerlund (2007) was carried out. The input parameters were remained the same as they were defined in the previous step, the model specification include only constant as a trend was shown statistically insignificant for all categories. The table 5 shows the results of the cointegration analysis including the value of the parameter α_i, which determines the speed at which the system corrects back to the equilibrium relationship after a sudden shock.

[5] Two companies with the headquarters' location in Mexico and Brazil were combined into one group
[6] With the help of Eviews 6 software

(0; 2) MBOE/D	***p-value***	***Robust p-value***	α_i
$G\tau$	0,000	0,000	
G_α	0,002	0,010	-0,4834
$P\tau$	0,181	0,230	
P_α	0,247	0,337	
(2; 4) MBOE/D			
$G\tau$	0,005	0,017	
G_α	0,000	0,003	-0,5932
$P\tau$	0,002	0,020	
P_α	0,000	0,013	
(4; 6) MBOE/D			
$G\tau$	0,501	0,417	
$G\alpha$	0,410	0,270	-
$P\tau$	0,851	0,843	
$P\alpha$	0,683	0,703	
(8; 10) MBOE/D			
$G\tau$	0,001	0,000	
$G\alpha$	0,007	0,013	-0,8522
$P\tau$	0,001	0,000	
$P\alpha$	0,000	0,013	

Table 5 Westerlund Error-correction-based panel cointegration test between revenues and oil price

The table above (table 5) indicates that in case of the 1st category with the lowest volume of oil production the cointegration relationship was evident on the 5% level of significant for at least one company, but not for the whole group. After the sudden shock the system corrects back to the equilibrium at the speed 0,48, which means that 48% of imbalance is corrected during one quarter. For the 2nd category there is a cointegration relationship for all companies on the 5% level of significance. Correcting back to the equilibrium is carrying out at the speed 0,59. For the companies from the 3rd category the cointegration relationship was statistically insignificant for all companies, so the value of α_i is not stated. For the last category, which includes the only one company with the largest volume of production the cointegration relationship was proved as statistically significant and 85% of disequilibrium is corrected during one quarter. So in all cases mentioned above the estimation of α_i is less than one and has a negative sign which corresponds to the econometric theory. It can be also noticeable that the volume of the oil and natural gas equivalents is a stronger control factor than the headquarters' or parent company's location, which was examined above. Based on the values of the α_i estimation in the individual categories it can be seen in the group with bigger volume of production the speed at which the system corrects back to the equilibrium is higher. Thus the companies with bigger overall volume of oil and natural gas equivalents' production are more stable towards sudden shocks in oil price. This conclusion should be considered with some limitations as firstly for the 3rd category the cointegration relationship was not proven, and secondly each category contains a different number of analyzed companies which could have its influence on the results as well.

In case of the analysis of the relationships between the logarithms of the revenues of the oil companies and gas price there are no such evident results. When the control variable is the headquarters' location the results were similar for two types of the gas price. The cointegration relationship was proven for the companies with headquarters' location in Russia with the speed correction back to the equilibrium above 0,8 (0,88 and 0,86). Such finding can be explained by the fact that Russia has the world's largest natural gas reserves, so the existence of the long term relationship is obvious. For all remained categories the null hypothesis of no cointegration was not rejected. Considering the next control variable in case of the price of U.S. natural gas exports the cointegration relationships exist only for the 3rd category (4-6 MBOE/D) for all companies from this category on the 10% level of significance, as well as, for the last category, which contains only one company with the overall production from the interval 8-10 MBOE/D but with $|\alpha_i| > 1$. If the gas price is modeled by Natural Gas spot price at the Henry Hub, the situation remains the similar with the difference that in the 3rd category the cointegration relationship was not proven. Although in most cases the null hypothesis of no cointegration was not rejected the analysis of this relationship remains very questionable and requires deeper analysis of the way of creation and specification of the gas price in the world considering geopolitical and strategic factors which is not the main of interest of this article.

5 Conclusion

The current paper examined if difference in control factors can cause different response of the biggest oil companies' revenues to the changes in the oil and gas price by using panel data tools. The analysis included two control variables: headquarters' or parent company's location and the volume of the oil production. Each control variable distinguishes four categories. In case of the 1st control variable the definite cointegration relationship between logarithms of the revenues and oil price was proven for the companies with headquarters' location in Russia and Europe. In case of USA the cointegration relationship was proven for at least one company. The speed at which the system corrects back to the equilibrium for Russia and USA had higher value than in case of Europe. In case of the second control factor referring to the production volume the cointegration relationship was proven for all categories excepting the category 4-6 MBOE/D. The interesting finding is that for the group with larger volume of production the speed at which the system corrects back to the equilibrium is higher.

Considering the same categories the cointegration analysis between the logarithms of the companies' revenues and gas price had not such clear results. In case of 1st control variable the long-term relationship was proven for the companies with headquarters' location in Russia. For all remained categories the cointegration relationship was insignificant. For the second control variable referring to the volume of the oil production the cointegretion relationship was proven for the last two categories spanning the volume more than 4 MBOE/D but for the gas price modeled by Natural Gas spot price at the Henry Hub the cointegration relationship for 3rd category was not proven. In case of the last category containing only one company (8, 1 MBOE/D), the long-term relationship was proven but $|\alpha_i| > 1$.

The limitation of this analysis consists in the fact that different categories contain different number of analyzed companies which could have its influence on the final results. The bigger number of analyzed companies could improve the accuracy of the conducted research. In these oil and gas issues the geopolitical and strategic factors should be taken into account.

Acknowledgements

I gratefully acknowledge the support of this project by the research grant VŠE IGA F4/58/2014, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] EBERHARD M.*; XTCD: Stata module to investigate Variable/Residual Cross-Section Dependence.* [online]. [cit. 2014-03-30]. http://fmwww.bc.edu/repec/bocode/x/xtcd.html.

[2] Index Mundi. *Henry Hub Natural Gas Futures End of Day Settlement Price*. [online]. [cit. 2014-04-27]. http://www.indexmundi.com/commodities/?commodity=natural-gas&months=360.

[3] LEWANDOWSKI, P.: PESCADF: *Stata module to perform Pesaran's CADF panel unit root test in presence of cross section dependence.* [online]. [cit. 2014-03-30]. http://fmwww.bc.edu/repec/bocode/p/pescadf.html.

[4] Persyn, D., Westerlund, J.: Error Correction Based cointegration Tests for Panel Data. *Stata Journal*. 8 (2). (2008), 232–241.

[5] Pesaran, M. Hashem: General Diagnostic Tests for Cross Section Dependence in Panels. *IZA Discussion Paper*. No 1240 (2004).

[6] Pesaran, M. Hashem: A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics.* 22(2). (2007), 265–312.

[7] U.S. Energy Information Administration. EUROPE Brent Spot Price. [Online]. [cit 2013-11-23]. http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RBRTE&f=M.

[8] U.S. Energy Information Administration. *Price of U.S. Natural Gas Exports.* [Online]. [cit 2014-02-20]. http://www.eia.gov/dnav/ng/hist/n9130us3m.htm.

[9] Westerlund, J. 2007. Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics* 69, 6, (2007): 709–748.

[10] Wikinvest.*Companies revenues.* [online]. [cit. 2014-02-20] http://www.wikinvest.com/site/About_Wikinvest.

[11] WTRG Economics. *Oil Price History and Analysis.* [Online]. [cit 2013-11-23]. http://www.wtrg.com/prices.htm.

Analysis of the dependence of GDP of the Russian Federation and the European Union on the Oil Demand and Oil Supply

Kuchina Elena[1], Tichá Michaela[2]

Abstract. This paper examines the dependence of GDP of the Russian Federation and the European Union on the World oil demand and supply. The present study focuses on the discovering the character of the relationships among the considering variables, on the determination and comparison the force of such dependence. Analysis is based on quarter data spanning the period from the 1st quarter of 2000 to the 3rd quarter of 2013.
Due to the nonstationary time series of the examined variables, the cointegration analysis was conducted. Based on it the vector error correction model (VEC) was chosen, which enables to discover a long-term as well as a short-term dynamic among considered variables. To determine the optimum lag length, as a basis for the cointegration analysis, a VAR model was constructed. At the end of the paper the short term dynamics in the concept of Granger causality was examined with the subsequent application of "Impulse-Response" analysis.

Keywords: Granger causality, cointegration analysis, "Impulse-Response" analysis, VAR model, VEC model.

JEL Classification: C51
AMS Classification: 91B84

1 Introduction

The oil industry always plays a significant role in the World economic situation. And the key indicator of the economic growth is the GDP. So the interesting question to examine is discovering the dependence of the GDP on the World oil industry, which is determined by the World oil demand and supply. In this paper the dependence of the GDP of the Russian Federation and the European Union will be examined.

The EU and Russia are among the top 20 largest oil producers in the World. The EU takes the 17th place in the World ranking of crude oil production (based on data from 2012), and Russia takes the 3rd place (based on data from 2013). In the list of the largest consumption of the refined petroleum products the EU takes the 2nd place (based on data from 2012) and Russia is on the 6th (based on data from 2012) [2].

According to the mentioned facts the hypothesis of the mutual relationships between the GDP of the chosen economics and the World oil demand and supply can be intuitively justified, which, however, should be examined by the methods of the statistical and econometric analysis.

The aim of this paper is to analyze the dependence of the GDP of the EU and Russia on the World oil demand and supply. The present study also discovers the character of this dependence, determine and compare the force of this dependence. The study also determines what variables precede the change of the other variables by testing Granger causality, and also "Impulse-Response" analysis was performed.

The article is structured into three parts., which are data set and the methodology determining the continuity of the steps, then the empirical modeling and main findings will be presented.

[1] *University of Economics, Prague, Econometrics dep., W Churchill Sq. 4, Czech Republic*, e-mail: xkuce10@vse.cz
[2] *University of Economics, Prague, Econometrics dep., W Churchill Sq. 4, Czech Republic*, e-mail: xticm11@vse.cz

2 Data Set

The analysis was based on the quarter data spanning the period from the 1st quarter of 2000 to the 3rd quarter of 2013. It was obtained 55 observations of the following macroeconomic variables:

- The GDP of the EU in billions euros (gdp_eu), The GDP of Russia in billions rubles (gdp_r) – data was obtained from Eurostat [3] and the financial information portal QUOTE.RBC.RU [8].
- The total world oil demand was calculates as a sum of OECD and non-OECD oil demand, and the same for the total world oil supply, which was obtained as a sum of the OECD, non-OECD and OPEC oil supply in thousands barrels per day (t_oil_d, t_oil_s) – International Energy Agency, Oil Markets Reports [6].

3 Methodology

In this section the procedure applied in the empirical part will be defined. The whole statistical and econometrical analysis was hold with the help of the software Eviews 6.

As the first step of the analysis the stationarity of the considered time series was tested by using Augmented Dickey-Fuller test (in the following text just ADF) applied to the logarithmic data including a constant and trend (the logarithmic data are used because of heteroscedasticity). Subsequently the order of integration was determined. Due to the nonstationarity of the examined variables and the absence of the precise theoretical economic definition between the considered variables the Vector error correction model (VEC) was chosen in this study. The general VEC model can be expressed as [5]:

$$\Delta X_t = \rho_1 Z_{t-1} + \delta_1 U_{t-1} + (\Delta X_t, \Delta Y_t)_{-i} + \varepsilon_{1t},$$
$$\Delta Y_t = \rho_2 Z_{t-1} + \delta_2 U_{t-1} + (\Delta X_t, \Delta Y_t)_{-i} + \varepsilon_{2t},$$

where a linear combination of the variables integrated by first order is stationary. Parameters δ_1 a δ_2 measure the speed of adjustment to equilibrium in time and should have a negative sign.

Before performing the VEC(p) model the corresponding cointegration analysis was taken by using Johansen cointegration test, under which Trace statistics and Max-Eig statistics were computed. There are several ways of the specification of the deterministic trend in data and in cointegration equation [4]:

- There is no deterministic trend in data, there is neither intercept nor trend in CE.
- There is no deterministic trend in data, there is intercept (no trend) in CE.
- There is deterministic trend in data, there is intercept (no trend) in CE.
- There is deterministic trend in data, there is intercept and trend in CE.
- There is deterministic quadratic trend in data, there is intercept and trend in CE.

The reasons for the choice of the particular specification will be discussed in more details in the empirical part of this study.

As input information for the cointegration analysis is the information about the optimal lag order, which was obtained based on the value of information criterions (Schwarz (SC) and Hannan-Quinn (HQ) information criterion) [5]. The values of these information criterions were stated based on the appropriate Vector autoregression model (VAR (p)). After performing the cointegration analysis and building the VEC model with the appropriate diagnostic tests, the Granger causality analysis was applied to examine the short-term dynamics with subsequent application of the "Impulse-Response" analysis.

The continuity of the steps, determined above, was applied to analyze the dependence of the GDP of Russia and the EU on the World oil demand and supply.

4 Empirical Modeling

Due to the logarithmic transformation the fluctuation in the values of the considered variables were reduced, but were not completely eliminated. So the problem of such significant fluctuation can be solved by dummy variables. Their necessity is proved by failure the residuals of the final model to meet the classical requirements.

Stationarity of the considered time series was tested by ADF unit root test, which was applied to logarithmic (including a constant and trend) time series. The following table (Table 1) summarizes the p-values of the ADF unit root test applied to the logarithmic data and their differences including a constant and trend.

	gdp_r	gdp_eu	t_oil_d	t_oil_s
log	0,9812	0,6697	0,4950	0,7628
dif_log	0,0000	0,0095	0,0000	0,0000

Table 1 ADF test

Based on the results in Table 1, the null hypothesis of the existence of the unit root cannot be rejected in case of all variables on the 1% level of significance. In case of their first differences the null hypothesis can be rejected on the 1% level of significant for all variables. So as a result of stationary analysis can be stated that all considered variables are I(1). This finding can be a starting point for the following advanced analysis.

4.1 Modeling of the multivariate non-stationary time series

Following the methodology chapter the first step is to determine the optimal lag length based on the information criteria's values obtained from the constructed VAR (*p*) model, which is a starting point for cointegration analysis. The next step is the cointegration analysis. Finally, after defining all baseline characteristics and confirmation the basic assumptions vector error correction model VEC (*p*) can be applied. Input endogenous variables are:

- Y_{Rt} – logarithm of the GDP of Russia (billions RUB), t = 1,2, 3, …, 55,
- Y_{EUt} – logarithm of the GDP of EU (millions EUR), t = 1,2, 3, …, 55,
- D_t – logarithm of the world oil demand (thousands barrel/den), t = 2, 3, …, 55,
- S_t – logarithm of the world oil supply (thousands barrel/day), t = 2, 3, …, 55.

Modeling the relationship between GDP of EU and total oil demand and supply

The optimal lag length based on the value of *SC* and *HQ* information criteria was chosen equal to one, i.e. $p = 1$. However, when testing the residua's properties, it was shown that either the hypothesis of normal distribution or serial independence assumption was not met on the 5% level of significance. This fact indicates that obtained VAR (1) cannot be considered as acceptable for the following cointegration analysis. As it was mentioned above, the problem of significant fluctuations can be removed by dummy variables. Two dummy variables were assigned to each variable, excepting the logarithm of EU GDP (based on its course line the only one fluctuation down was significant (in 1Q2009), so only one dummy variable was assigned to it). The first variable takes the value one if there are fluctuations down, and in other cases the value is zero. In case of the second dummy variable the value one indicates fluctuations upward. After inclusion dummy variable as exogenous variables, VAR (1) model was repeatedly applied. The values of the information criterions indicated the optimal lag still equal to one. However, residual tests did not show acceptable results. The dummy variable corresponding to the fluctuations upward in case of logarithm of the oil demand was not statistically significant (P-value>0,05), so it was removed. Significant fluctuations of the residual graph were solved by dummy variables as well. Diagnostic residual tests didn't reject that the basic diagnostic residuals assumptions were met in terms of normality, serial independence and homoscedasticity on the 5% level of significance, and besides it the values of the coefficient of determination were above 97% for all three equations, so the obtained VAR (1) model can be acceptable for the following cointegration analysis

Cointegration analysis was performed with the help of Johansen test, based on which the one cointegration vector was confirmed for the first and second case (*see Methodology*). First case can be used only if all series have zero mean, and the assumption of using the second case is the fact that none of the series has a trend. Both these preconditions are not met in this problem, so the conclusion of no cointegration can be made. Thus the VAR analysis of first-difference series can be proceed with Granger causality analysis. According to results gained from the VAR (2) model it was shown that in a short term ΔY_{EUt} depends on $\Delta Y_{EU,t-1}$ and ΔD_{t-1}. In case of ΔD_t in a short term its dependence was statistically significant on ΔD_{t-1} and $\Delta D_{t-2.}$ Finally, the independence of ΔS_t on any of considering variables was not rejected as statistically significant on the 5% level of significance. The independence on the macroeconomic changes, which influence was inbuilt in dummy variables, was rejected in all equations, which is intuitively expected result. The coefficients of determination in all equations are relatively low (0,66; 0,57; 0,31) that can indicate that there is no strong dependence of EU GDP on the world oil depend and supply, and in order to improve the model's quality many other factors should be taken into account.

As a next step the Granger causality was tested. The null hypothesis tests if one variable does not Granger-causes other variable. In case of rejection of the null hypothesis the lagged value of the first variable contributes to increase the prediction's accuracy of other variable [5]. Since all p-values are greater than 5% (table 2), we cannot reject the null hypothesis and thus there is no Granger-causality between EU GDP and the world oil demand.

Dependent variable	ΔY_{EUt}		ΔD_t		ΔS_t
ΔD_t	0,1150	ΔY_{EUt}	0,5335	ΔY_{EUt}	0,9009
ΔS_t	0,0848	ΔS_t	0, 1512	ΔD_t	0,5656
All	0,1058	**All**	0,1596	**All**	0,7452

Table 2 Granger causality test (*p*-values)

Modeling the relationship between GDP of Russia and total oil demand and supply

Similar to the procedure applied above, firstly, the optimal lag was defined equal to one based on the information criterion (*SC*), then the compliance with basic residual assumptions was tested. The problem of autocorrelation occurred on the 5% level of significance. Analogous to the previous model dummy variables were included to solve the problem of significant fluctuations. Two dummy variables corresponding to the fluctuations up and down were assigned to each considering variable plus two dummy variables corresponding to extreme fluctuations upward and down based on the residual graph, which can be caused by lagged reaction to the so called oil shocks when the oil price achieved extremely high or extremely low values. Then VAR (1) model was reputedly applied. Basic assumptions on residuals were met on the 5% level of significance.[3] In case of logarithms of Russian GDP and oil supply the one of two dummy variables (in both cases dummy corresponding to the fluctuations upward) were not statistically significant, so they were excluded from the model. The value of *SC* information criterion indicates that the same optimal lag length equal to one. So the gained model can be considered as appropriate for the cointegration analysis.

Based on Johansen test the existence of cointegration vector was confirmed in the first three cases (see *Methodology*). As it was explained above first two cases are not appropriate for this model, so the third case considering a deterministic trend in data and intercept (no trend) in CE was chosen. Then VEC (1) model was applied. When the residual diagnostic tests did not reject the meeting of all basic assumptions on the 5% level of significance the interpretation step can be followed. Coefficient of determination in the equations corresponding to Y_{Rt} and D_t are above 84%, what indicates relatively good model fitting with data. In case of the equation corresponding to S_t, the coefficient of determination is equal only to 52%, but the purpose of this paper consists in the analyzing of the existing relationships between examined variables not in forecasting, so the estimated model can be considered appropriate for the following analysis. The coefficient of the cointegration equation was statistically significant for the equations corresponding to Y_{Rt} and D_t, but for the second equation it has a positive sign, what is contrary to the econometric theory [5]. In Y_{Rt} the system corrects back to the equilibrium relationship after a sudden shock at the speed equal to 0,19 during one quarter (the parameter estimation is -0,19, so it is less than one and has a negative sign corresponding to the econometric theory). On closer examination of the cointegration equation including Y_{Rt}, D_t, S_t and a constant, it can be noticeable that estimations of all parameters were statistically significant on the 5% level of significance. Short term dynamics was statistically significant between ΔY_{Rt} and ΔD_{t-1}. In case of ΔD_t there was not proven short term dependence on the difference of Y_{Rt} and S_t. However, in case of the last equation corresponding to ΔS_t, it was proven that it depends in a short term on ΔY_{Rt-1}, which is in conformity with the intuitive hypothesis, as Russia plays a significant role in the world oil production. It is also interesting to notice that in case of all three equations corresponding to Y_{Rt}, D_t, S_t the short term dependence on the dummy variables describing the extreme fluctuation of the residual graph was proven as statistically significant. So, the role of macroeconomic factors in the oil industry is vast.

After determining the character of the relationships between the considering variables the next step consisting in the testing of Granger causality can be followed. Testing the Granger causality in the integrated and cointegrated systems is applied on the differential time series, i.e. I(0). *p*-values of the Granger causality testing are in the following table (table 3).

Dependent variable	ΔY_{Rt}		ΔD_t		ΔS_t
ΔD_t	0,0012	ΔY_{Rt}	0,0496	ΔY_{Rt}	0,0280
ΔS_t	0,8717	ΔS_t	0, 1527	ΔD_t	0,4172
All	0,0005	**All**	0,0629	**All**	0,0894

Table 3 Granger causality test (*p*-values)

[3] The null hypothesis of serial independence in case of the first lag can be accepted on the 10% level of significance.

Based on the results in Table 3 it can be noticed that ΔD_t Granger causes ΔY_{Rt} on the 5% level of significant, in case ΔS_t the causal effect to ΔY_{Rt} was not proven but both ΔD_t and ΔS_t Granger-cause ΔY_{Rt}. When the dependent variable is ΔD_t, ΔY_{Rt} Granger-causes it but because of *p*-value is very closed to the 5% level of significant it remains questionable. Finally, when ΔS_t is in the role of dependent variable, it can be seen that ΔY_{Rt} Granger-causes it on the 5% level of significant. So the results of the investigation of Granger causality can summarized as:

$\Delta D_t \rightarrow \Delta Y_{Rt}$
$\Delta Y_{Rt} \rightarrow \Delta S_t$

After analyzing the causality in Granger concept it can be moved to "Impulse-Response" analysis allowing to examine the response of one variable caused by impulse in other one in multivariate system [1]. Under the "Impulse-Response" analysis it was examined the response of Y_{Rt} to the exogenous shock in D_t and the response of S_t to the exogenous impulse in Y_{Rt}. Due to unavailability of Eviews 6 to build confidence intervals for VEC model the JMulTi software [4] was used allowing to build bootstrapped confidence intervals [7]. Based on the results shown on Figure 1 it can be concluded that Y_{Rt} reacts to the exogenous shock in D_t by immediate fluctuating down then in the 1st quarter there is a turning point when the fluctuation changes its direction by opposite side, after 4th quarter the stabilized on the same level. Analyzing the reaction of S_t to the impulse in Y_{Rt} it can be noticed that there is immediate fluctuation upward but after 1st quarter it steadies on the approximately the same level.

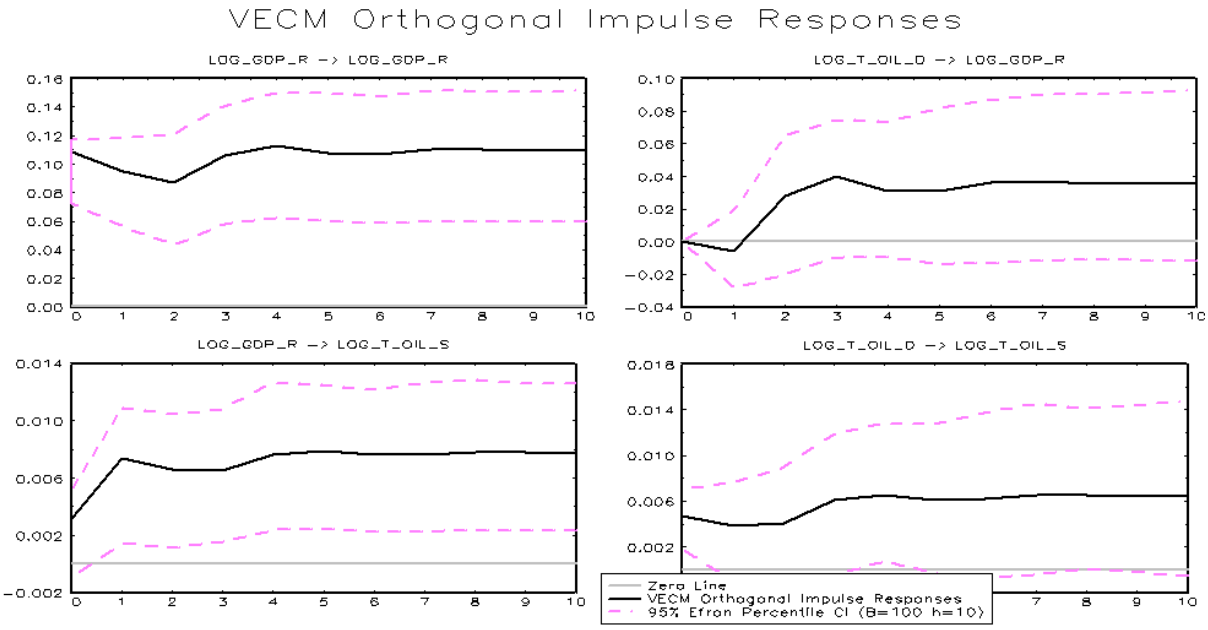


Figure 1 "Impulse-Response" Function

5 Conclusion

The current paper has constituted the econometric analysis of the dependence of the GDP of Russia and the EU on the world oil demand and supply. Due to the fact that the EU and Russia are among the top 20 largest oil producers in the World, it was assumed that there can be the mutual relationships between the GDP of the chosen economics and the world oil demand and supply. To analyze such relationships between the considered variables the VEC model was chosen.

In case of modeling relationships between the EU GDP and world oil demand and supply the conclusion of no cointegration was done. Even though the cointegration vector was proven for the first cases of the CE specification, their assumptions were not met. So, because of the long term equilibrium was not proven the analysis of the

[4] http://www.jmulti.de/ Time series analysis with Java

short-term dynamics by using VAR model (applied on the first differenced variable) followed. The quality of the model was relatively low based on the coefficient of determination, as well as, the Granger causality analysis did not reject the null hypothesis that any variable Granger causes the other variables on the 5% level of significant. So this fact indicates that for the appropriate modeling other factors should be taken into account.

In case of the GDP of Russia the existence of the long-term equilibrium between the Russian GDP and world oil demand and supply was proven In case of the logarithm of the Russian GDP the system corrects back to the equilibrium relationship after a sudden shock at the speed equal to 0,19 during one quarter. In a short-term the logarithm of the GDP of Russia depends only on its differenced value lagged by first order. After application of the Granger causality concept it was noticeable that the difference of the logarithm of the oil demand Granger-causes the difference of the logarithm of the Russian GDP, which Granger causes the difference of the logarithm of the oil supply. This finding corresponds to the economic verification as Russia is among the largest oil producers in the World. The results of the "Impulse-Response" analysis indicate that the reaction of the logarithm of the Russian GDP to the exogenous shocks in logarithm of the oil demand, as well as the response of the logarithm of the oil supply to impulse in the logarithm of the Russian GDP stabilized rather soon, in first case after 4^{th} quarter and in the second case after 1^{st} quarter.

Finally, it is important to emphasis in the analysis of the relationships between the GDP of the chosen countries and the world oil demand and supply it is evident that the GDP cannot be determined only by oil industry even though it plays a significant role. Many other macroeconomic and geopolitical factors should be taken into account.

Acknowledgements

We gratefully acknowledge the support of this project by the research grant No. VŠE IGA F4/24/2014, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] ARLT, J., ARLTOVÁ, M.: *Ekonomické časové řady*.Professional Publishing, Praha, 2009
[2] Central Intelligence Agency. The World Factbook. *Russia*. [online]. [cit. 2014-04-07]. https://www.cia.gov/library/publications/the-world-factbook/geos/rs.html.
[3] Eurostat. [online]. [cit. 2014-03-03] http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=teina010&plugin=1
[4] EViews 6: *EViews 6 User's Guide II*. USA, Quantitative Micro Software, LLC, 2007
[5] HUŠEK, R.: Ekonometrická analýza. Oeconomica, Praha, 2007.
[6] International Energy Agency. Oil Market Reports. [online]. [cit. 2014-03-03]. http://omrpublic.iea.org/tablearchivesearchres.asp?select5=%25&Submit222=Submit.
[7] LÜTKEPOHL H., KRÄTZIG M.: *Applied Time series Econometrics*. Cambridge University Press 2004. [online]. [cit. 2014-05-10]. http://f3.tiera.ru/2/G_Economics/GL_Econometrics/Lutkepohl%20H.,%20Kratzig%20M.%20(eds.)%20Applied%20time%20series%20econometrics%20(CUP,%202004)(ISBN%200521547873)(350s)_GL_.pdf.
[8] QUOTE.RBC.RU. *GDP of Russia*. [Online]. [cit 2014-03-03]. http://quote.rbc.ru/macro/indicator/1/2.shtml

Comparison of prediction power of BVAR and DSGE models

Petra Kummerová[1], Miroslav Hloušek[2]

Abstract. This paper analyses macroeconomic development in the Czech Republic using structural steady state Bayesian VAR model, classical VAR model, and New Keynesian DSGE model for a small open economy. The steady state BVAR was developed by Villani (2009), DSGE model is borrowed from Justiniano and Preston (2004). Foreign sector in VAR models is implemented as block-exogenous. The models are estimated on data of Czech economy covering period 1996Q3 – 2013Q3. All models are compared on the basis of prediction performance measured by RMSE, impulse responses and variance decomposition. The results show that steady state BVAR has better prediction power for longer time horizons which is in accordance with its structure. On the other hand, DSGE model better predicts behaviour of variables for short periods, up to four quarters. According to RMSE computed for different horizons over the whole dataset, classical VAR model predictions were the least accurate in the majority of cases, DSGE model have the best predictive power for behaviour of foreign variables, and steady state BVAR overcomes the other models in prediction of interest rates in domestic economy.
Keywords: steady state BVAR model, New Keynesian DSGE model, forecasting performance.

JEL Classification: E37
AMS Classification: 62P20

1 Introduction

The interest in macroeconomic analysis started with J.M. Keynes at the beginning of the 20th century. Nevertheless, in order to conduct effective macroeconomic policy the agents need timely and accurate information about the current and future economic development. Later on, various macroeconomic models have been developed to accommodate these needs. This paper therefore focuses on the accuracy of predictions of two the most frequently used types of models in macroeconomic analysis. The first model is Bayesian VAR, whereas the second one is New Keynesian DSGE model. The Bayesian VAR model is used in a slightly amended version which allows implementation of informative priors on the steady state values or the long term rates of growth. The DSGE model is borrowed from Justiniano and Preston (2004). The models are estimated on Czech data and are compared in terms of prediction power.

The rest of the paper is organized as follows. Section 2 deals with the methodologies of both models. Section 3 describes data used in the estimation and the empirical implementation of them. Section 4 compares the estimated results and predictive performance of both models, the last section concludes.

2 Models

In order to be able to compare the results obtained from the underlying models, both of them are estimated using Bayesian techniques and the same variables are used.

2.1 Methodology of the steady state BVAR

The steady state BVAR was developed by Villani (2009) and has the following form:

$$\Pi(L)(x_t - \psi d_t) = \varepsilon_t \tag{1}$$

where $\Pi(L)$ is the polynomial matrix of the lag operator with lag p which could be rewritten as $\Pi(L) = \mathrm{I} - \Pi_1 L - \ldots - \Pi_p L^p$, d_t is the vector of deterministic trends[3], ε_t is iid[4] vector of errors with zero mean, $E(\varepsilon_t) = 0$, and the covariance matrix $E(\varepsilon_t \varepsilon_t') = \Sigma$, the term $(x_t - \psi d_t)$ stands for the variations from the long run trend, ψ

[1] Masaryk University, Department of Economics, Lipová 41a, 602 00 Brno, p.kummerova@seznam.cz.
[2] Masaryk University, Department of Economics, Lipová 41a, 602 00 Brno, hlousek@econ.muni.cz.
[3] If the BVAR contains only constant , $d_t$is a vector of ones, if the model contains also time trend, structural shift or seasonality, $d_t$is a matrix.
[4] Identically and independently distributed.

are estimated parameters of deterministic trends (d_t). If the variables are in levels, the variations from the long run trend mean variations from the steady state values. If the variables are in growth rates, the variations from the long run trend mean variations from the long run growth rates.

Analogously to the reduced form, the structural form can be derived. In order to analyze the structural shocks and due to better comparability with the DSGE model, the structural form is used in this paper.

The steady state form enables to incorporate our prior beliefs about the vectors Π and ψ which are assumed to come from the normal distribution:

$$\Pi \sim N_{pn^2}(\theta_\Pi, \Omega_\Pi) \tag{2}$$

$$\psi \sim N_n(\theta_\psi, \Omega_\psi) \tag{3}$$

The prior values of the lag coefficients lying out of the diagonal of the matrix Π are set to be zero. The values on the diagonal corresponds to our prior beliefs about the first own lag. When the structural form is estimated, we assume that the structural shocks ε_t are *iid* with zero mean value and identity covariance matrix. The structural coefficients in the intratemporal matrix are obtained through Cholesky identification. The restrictions on the matrix of coefficients have the form of lower triangular matrix. Thus, the variables can affect only the successive variables in the model and the coefficients are estimated recursively.

The reduction parameters are set according to the Litterman prior as follows:
$\lambda_1 = 0.2$ overall tightness$\lambda_2 = 0.5$ cross-equation shrinkage
$\lambda_3 = 1$ lag decay parameter
$\lambda_4 = 0.001$ erogeneity tightness – through this parameter the block erogeneity is modeled.

The model is applied to the Czech Republic which is considered as a small open economy. Therefore, the euro area is assumed as the foreign economy that affects the domestic one and not vice versa. This assumption is implemented through the erogeneity tightness parameter which is set as sufficiently low.

The reduction parameters enter the model in the following form:

$$S\left(\pi_{ij}^{(k)}\right) = \begin{cases} \dfrac{\lambda_1 s_i}{k^{\lambda_3} s_j} & \text{if } i = j \text{ (own lag)} \\ \dfrac{\lambda_1 \lambda_2 s_i}{k^{\lambda_3} s_j} & \text{if } i \neq j \text{ (cross} - \text{equation lag)} \end{cases} \tag{4}$$

Where $\pi_{ij}^{(k)}$ is the k^{th} lag coefficient for j^{th} variable in equation for i^{th} variable, s_j and s_i are the corresponding standard deviations[5].

The block erogeneity enters the model as follows:

$$S\left(\pi_{ij}^{(k)}\right) = \frac{\lambda_1 \lambda_2 \lambda_4 s_i}{k^{\lambda_3} s_j} \tag{5}$$

The whole model is estimated through Gibbs sampler that iteratively takes samples from the conditional posterior densities. The joint probability density is subsequently derived from these samples. The number of iterations has been set to 10 000 and the first 5 % of samples has been discarded. The posterior distribution is derived from the rest of the samples.

2.2 Methodology of the New Keynesian DSGE model

The DSGE model of Justiniano and Preston (2004) has been chosen with respect to the structure of the Czech economy that is a small open economy influenced by the development abroad. The euro area had been chosen as the foreign economy. Here, the model is described only verbally; for details see the original paper. The model is specified through the theoretical equations describing the behavior of domestic and foreign households, domestic and foreign producers, domestic retail firms importing foreign goods and the market clearing conditions including uncovered interest rate parity.

The maximization problem of the households is quite standard. The households make decisions about consumption, supply of work, and holding of bonds. The consumption is affected by the habit formation and consists of consumption of both domestic and foreign goods. All their decisions are restricted by the budget constraint.

[5] $S\left(\pi_{ij}^{(k)}\right)$ is the matrix of standard deviations.

It is assumed that the producers' market consists of a continuum of monopolistically competitive producers. All the producers set their prices according to the Calvo (1983) style pricing which means that part of them set the prices optimally in each period *t*, and part of them only adjust the prices through indexation to the past inflation. All the producers maximize the present value of their future income subject to the restrictions in form of the demand function. Analogously, the pricing carried out by retail firms importing the foreign goods can be derived. The importers reset the price of imported goods in the domestic currency. Thus, the exchange rate is also part of their optimization problem.

All the equations are log-linearized around the steady state using the Taylor's series approximation. The estimation is carried out through Bayesian techniques. The Metropolis-Hastings algorithm is used as a posterior simulator. The number of samples from posterior density is 100 000 and the first 60% of the samples have been discarded. The Matlab function *csminwel* from Chris Sims is used as an optimization method.

3 Data and empirical implementation

Both of the models are specified through the following variables:

$$\boldsymbol{x}_t = (\, GDPea_t \; Pea_t \;\; IRea_t \;\; GDPcz_t \; Pcz_t \;\; IRcz_t \; RER_t \,)' \tag{6}$$

where $GDPea_t$ stands for demeaned quarterly growth rate of GDP of the euro area, Pea_t stands for demeaned quarterly inflation rate in euro area derived from HICP, $IRea_t$ stands for HP-filtered 3-months interest rates in the euro area, $GDPcz_t$ stands for demeaned quarterly growth rate of GDP of the Czech Republic, Pcz_t stands for demeaned quarterly inflation rate in the Czech Republic derived from HICP, $IRcz_t$ stands for HP-filtered 3-months interest rates in the Czech Republic, and RER_t stands for demeaned quarterly growth rate (change) of the real exchange rate.

Data cover the period 1996Q3-2013Q3 and the source is Eurostat. All the series are seasonally adjusted. The euro area is modeled as EA-17[6].and the inflation and GDP data were calculated backwards by Eurostat as aggregates for all 17 countries. In calculation of the euro area interest rates, the changing composition is used (EA11-2000, EA12-2006, EA13-2007, EA15-2008, EA16-2010, and EA17). The changes to the previous period are modeled with respect to the previous quarter $\Delta x_{t,t-1}$. The lag length in the BVAR is set to two.

3.1 Prior information

The priors on first own lag were set to 0.9 for all variables except of the real exchange rate where it was set to zero. The prior distribution of deviations from the long run values is presented in Table I. Due to the worldwide economic crisis an outlier variable was implemented for the period 2008Q3-2009Q3. Prior distribution of the outlier is also presented in Table 1. All priors are set according to the authors' view with reference to data.

The priors for DSGE model are mostly borrowed from Justiniano and Preston (2004 and 2006) or set according to data from Czech national accounts.

Variable	Lower bound of Constant	Upper bound of Constant	Lower bound of Outlier	Upper bound of Outlier
GDPea	0.25	1	-1	-0.25
Pea	0.1	0.75	-0.75	-0.1
IRea	0.1	0.75	-0.75	-0.1
GDPcz	0.25	1	-1	-0.25
Pcz	0.25	1.2	-1.2	-0.25
IRcz	0.25	1.2	-1.2	-0.25
RER	1	8	-8	-1

Table 1 Prior information in BVAR in % (q-on-q change)

4 Comparison of results of BVAR and DSGE

Model comparison is carried out through RMSE, impulse response functions and variance decomposition. Table 2 shows RMSE prediction errors for BVAR, DSGE, and maximum likelihood VAR model calculated

[6] In 2013, there were 17 countries in the euro area: Belgium, Estonia, Finland, France, Ireland, Italy, Cyprus, Luxembourg, Malta, Germany, the Netherlands, Portugal, Austria, Greece, Slovakia, Slovenia and Spain.

throughout the whole sample (1996Q3-2013Q3). The RMSE for BVAR is usually smaller for longer horizons than in the short run (upper panel). The opposite is true for DSGE model – with longer horizon the prediction error usually increases. Moreover, the classical VAR estimated by maximal likelihood method was tested as a benchmark. Nevertheless, the classical VAR model was the least accurate in most of the cases. The poor results of classical VAR could be seen particularly in the predictions for longer horizons. Despite the correctly predicted direction of change in macroeconomic variables during the worldwide crisis, none of the models was able to predict the full magnitude of it.

Horizon	GDPea	Pea	IRea	GDPcz	Pcz	IRcz	RER
1Q	0.66	0.25	0.42	0.88	0.36	0.33	2.94
2Q	0.73	0.36	0.72	0.99	0.48	0.5	2.97
3Q	0.81	0.41	0.95	1.07	0.61	0.68	3.16
4Q	0.86	0.41	1.1	1.14	0.69	0.83	3.24
8Q	0.39	0.2	1.18	0.52	0.5	1.11	1.32
12Q	0.35	0.16	0.55	0.52	0.46	1	1.19
Horizon	GDPeu	Pea	IRea	GDPcz	Pcz	IRcz	RER
1Q	0.48	0.14	0.29	0.76	0.36	1.17	2.6
2Q	0.53	0.22	0.49	0.9	0.54	1.49	3.1
3Q	0.49	0.26	0.66	0.81	0.63	1.72	3.11
4Q	0.49	0.28	0.83	0.71	0.69	1.68	2.92
8Q	0.63	0.27	1.18	0.87	0.6	1.73	2.7
12Q	0.67	0.24	1.15	0.88	0.51	1.33	2.49
Horizont	GDPea	Pea	IRea	GDPcz	Pcz	IRcz	RER
1Q	0.81	0.23	0.59	0.98	0.59	0.94	3.03
2Q	0.88	0.37	1.18	1.25	0.78	1.18	3.45
3Q	0.66	0.39	1.61	1.12	0.56	1.31	2.83
4Q	0.88	0.32	1.66	1.04	0.48	0.99	3.59
8Q	0.83	0.2	1.44	0.99	0.68	2.08	3.06
12Q	1.15	0.38	2.35	0.94	1.29	1.6	4.91

Table 2 RMSE prediction errors: BVAR (first), DSGE (second), and MLE VAR (third)

Direct comparison of prediction performance of BVAR and DSGE is shown in Table 3. The numbers are calculated from RMSEs as: (DSGE-BVAR)/BVAR and expressed in %. The most distinct differences are highlighted – red values indicate better prediction performance of BVAR, green values favors DSGE model. The results confirm previous findings: the DSGE model is more accurate in short run predictions up to four quarters. Only for prediction of domestic interest rate, DSGE model failed at all horizons. On the other hand, steady state BVAR is more accurate in long run predictions. That is in compliance with the model's methodology that incorporates the long term growth rates as informative priors. Generally, the DSGE model predicts the behavior of foreign variables with more accuracy.

Horizon	GDPea	Pea	IRea	GDPcz	Pcz	IRcz	RER
1Q	-27.4	-45.8	-30.5	-14.2	-0.2	256.8	-11.6
2Q	-28.2	-37.5	-31.6	-9.1	11.3	198.2	4.4
3Q	-40.1	-35.6	-30.9	-24.7	3.3	155	-1.6
4Q	-42.9	-31.2	-24.5	-37.2	0.8	102.5	-9.9
8Q	60.5	36.6	-0.2	68.4	19.4	55.5	104.6
12Q	90.6	50.8	110.9	68.2	10.1	33.3	108.8

Table 3 RMSE prediction errors: DSGE compared to BVAR in %[7]

Additionally, Figure 1 presents comparison of the cascade plots of BVAR and DSGE that depicts the iterative predictions. Figure 2 displays the cascade plot of classical VAR as a benchmark. The first prediction has been made from 2008Q1 and the length of each iterative prediction is 12 periods. It is worth mentioning that the DSGE model highly underestimated the development of interest rates in both regions around the year 2010. On the other hand, DSGE predicted the behavior of Czech inflation with more accuracy around the same period. The classical VAR is not very accurate in any of the periods which corresponds with the RMSE results.

[7] (DSGE-BVAR)/BVAR * 100

Regarding the impulses response functions,[8] the variables in both models reacted very similarly. The differences between the impulse responses were only in magnitude and longitude at some reactions. The similar findings apply for variance decomposition comparison.[9]

Moreover, the impulse response analysis showed that the reaction of domestic variables to the domestic shocks is usually with higher both magnitude and longitude in DSGE compared to BVAR. On the other hand, the reaction of domestic variables to the foreign shocks is usually higher and lasts longer in BVAR.

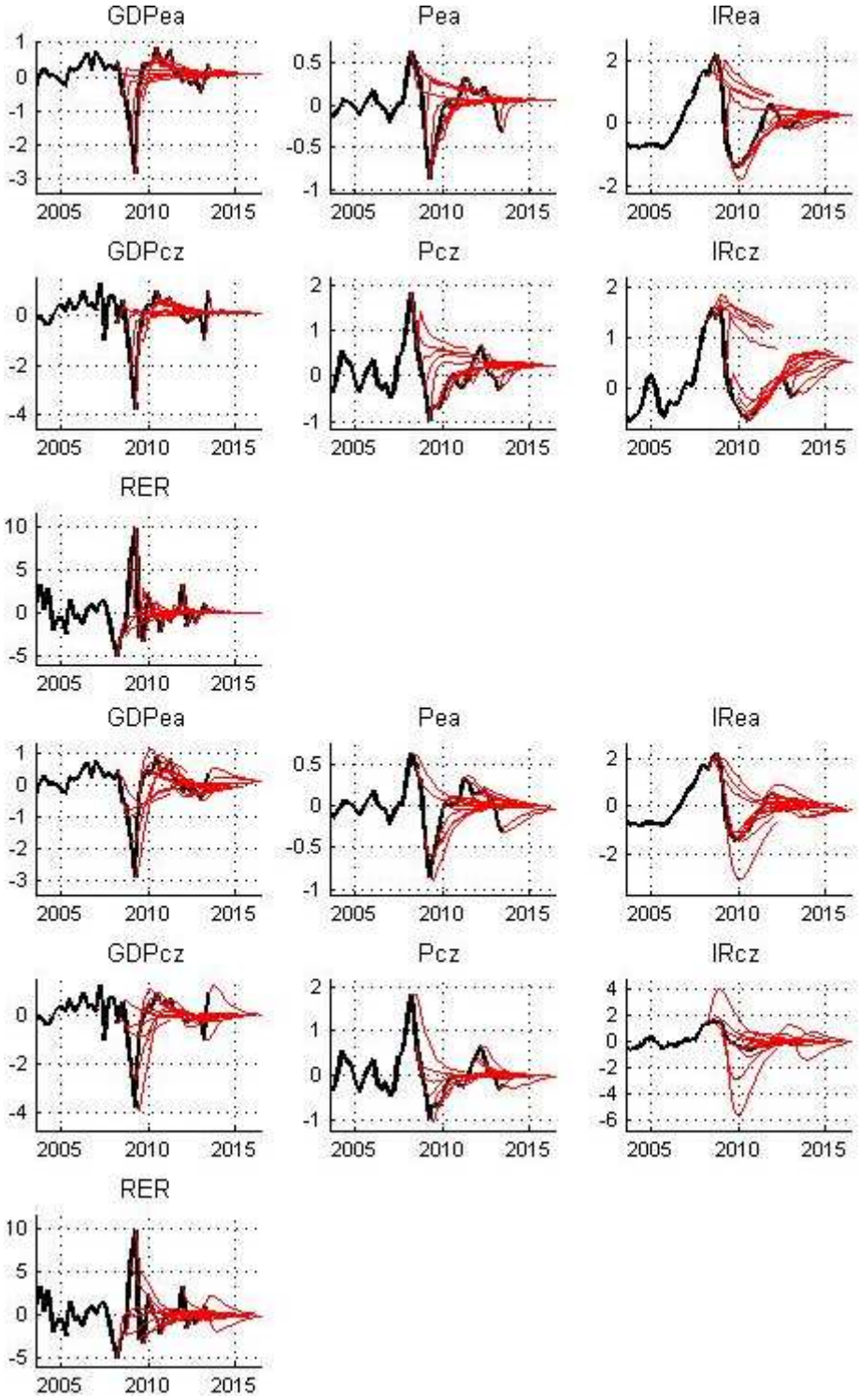


Figure 1 Cascade plot for BVAR (first in order) and DSGE (second in order)

[8] Not shown here.
[9] All these results are available upon request.

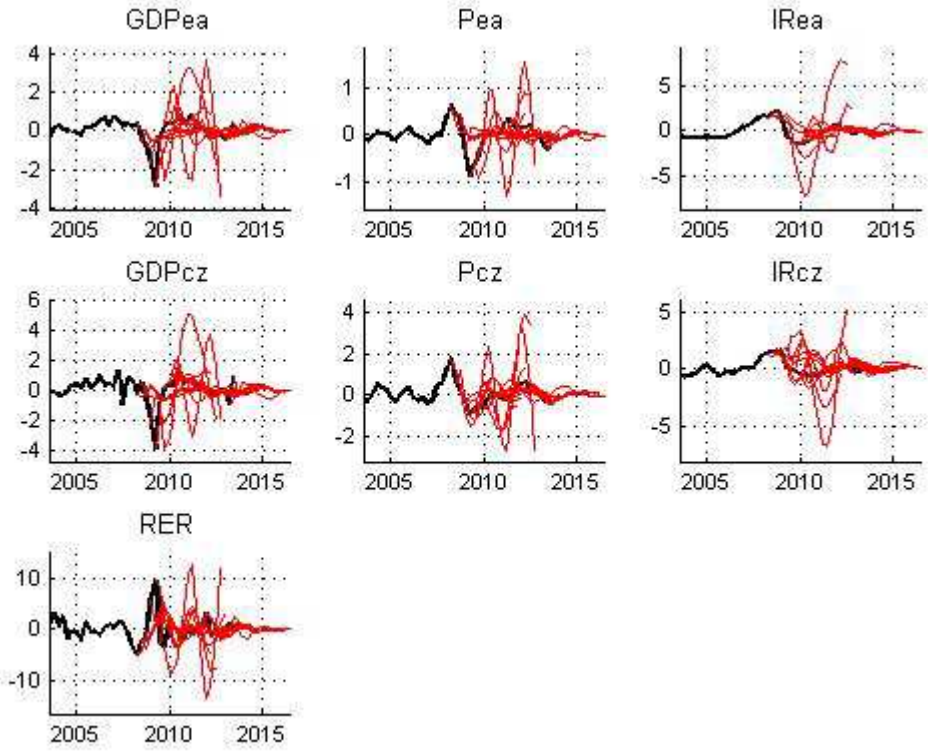


Figure 2 Cascade plot for MLE VAR

5 Conclusion

In this paper we have compared prediction performance of steady state BVAR, DSGE model and classical VAR model. The results show that none of the models is the most accurate in all cases which corresponds also with other studies on the similar topic. DSGE model is more accurate in short term predictions, whereas BVAR is more accurate in long term predictions. MLE VAR model was at almost all cases the least accurate. As the period used for the analyses covers also the worldwide economic crisis, the prediction performance might be underestimated, especially in BVAR as it uses the long run trend as prior information. It might be the case that the prediction performance of steady state BVAR in non-crisis period would be much more accurate.

Acknowledgements

This paper is supported by specific research project no. MUNI/A/0775/2013 at Masaryk University.

References

[1] Heij, C. *Econometric methods with applications in business and economics*. New York: Oxford University Press, 2004, xxv, 787 p. ISBN 01-992-6801-0.
[2] Calvo, G. Staggered prices in a utility-maximizing framework, *Journal of Monetary Economics 12*, 383—398, 1983.
[3] Justiniano, A., and Preston, B.: Monetary policy and uncertainty in an empirical small open-economy model. *Journal of Applied Econometrics*. 2010, vol. 25, issue 1, 93-128.
[4] Justiniano, A., and Preston, B.: *Small Open Economy DSGE models: Specification, Estimation and Model Fit*. Working Paper, Department of Economics, Columbia University, 2004.
[5] Villani, M.: *Inference in Vector Autoregressive Models with an Informative Prior on the Steady State*. Working Paper Series No. 181, Sveriges Riksbank, 2005.
[6] Villani, M.: *Steady State Priors for Vector Autoregressions*. Sweden, Research Division, Sveriges Riksbank, 2009.

Analysis of the efficiency of the electricity supplier selection depending upon the price changes

Martina Kuncová[1], Jana Sekničková[2]

Abstract. Electricity belongs to a very important commodities as its consumption is still rising. The transformation of the electricity market in the Czech Republic has led to the increasing number of suppliers offering this commodity to households and companies. The Czech Republic is divided into three network parts operated by three distributors so the households cannot choose the distributor but only the supplier. Each household has its tariff rate according to the supplier's conditions. The complete list of suppliers and their tariffs and prices is not providing an easy survey. It is influenced by the region where we choose the suppliers, by the amount of the electricity consumption and by electricity prices. Prices can change every year according to the regulations of the Energy Regulatory Office and also with respect to the decisions of each supplier. The price changes have an important impact on the final consumers. The selection of the suppliers is dependent upon the price development but it can be limited by the contract conditions of the previous supplier. The aim of this paper is the comparison of the suppliers based on the price changes during last 4 years via simulation model.

Keywords: Electricity network, Suppliers, Prices development, Simulation model.

JEL Classification: C44, C63, O13
AMS Classification: 90C15

1 Introduction

The economic development and changes on the markets force the suppliers and customers to study the situation, to speculate and to react on the changes. The electricity market of the Czech Republic is one of the markets that was specific till 2002 when the transformation process oriented at the fully liberalization has started. In the first phase (till 2006) only the companies could enter the market but since 2006 also the households can choose the electricity supplier on the retail market. The specificity of this commodity influences the number of the subjects of the market as there are not only suppliers and customers but also distributors, the Energy Regulatory Office (ERU) and the Operator of the market (OTE). OTE predicates the whole market consumption and analyses the differences, ERU regulates the prices of the transfer and distribution of the electricity. The increasing number of suppliers on the retail market embarrasses the position of the households as it is hard to follow the rules and the price changes on the market and so it is hard to choose the best (cheapest) product. Methods of multi-criteria evaluation of alternatives or simulation models can be used to solve this problem. In this article we try to describe the situation on the retail market in the period 2011-2014 with respect to the number of suppliers and its products and prices, and we compare the annual costs of the electricity for the given household when the monthly consumption is simulated.

2 Czech Electricity Market

Czech Republic is divided into three parts operated by three distributors (PRE, E.ON, CEZ). The customers cannot choose the distributor but only the supplier. The possibility for the change has started in 2006 when the liberalization of the Czech electricity market begun. In the first years the competition of the suppliers in the segment of medium-sized customers started to grow and also since 2009 the growing number of suppliers can be seen in the segment of small customers and households. In the period 2011-2013 we can see rapid increase in number of suppliers and the products they offer but in 2014 this trend does not continue (Figure 1). Especially households have started to take notice about the possibility to change the supplier to decrease its electricity expenses and so the boom in the switches has arose (Figure 2). Each household has its tariff rate according to the supplier's conditions. Various tariff rates are offered to households but their selection is limited as there are

[1] University of Economics, Prague, Department of Econometrics, W.Churchill Sq. 4, 130, 67 Prague 3, Czech Republic, e-mail: martina.kuncova@vse.cz
[2] University of Economics, Prague, Department of Econometrics, W.Churchill Sq. 4, 130, 67 Prague 3, Czech Republic, e-mail: jana.seknickova@vse.cz

limitations for the tariff usage. The complete list of suppliers and their tariffs and prices is not providing an easy survey. It is influenced by the region where we choose the suppliers and also by the electricity take-off amount.

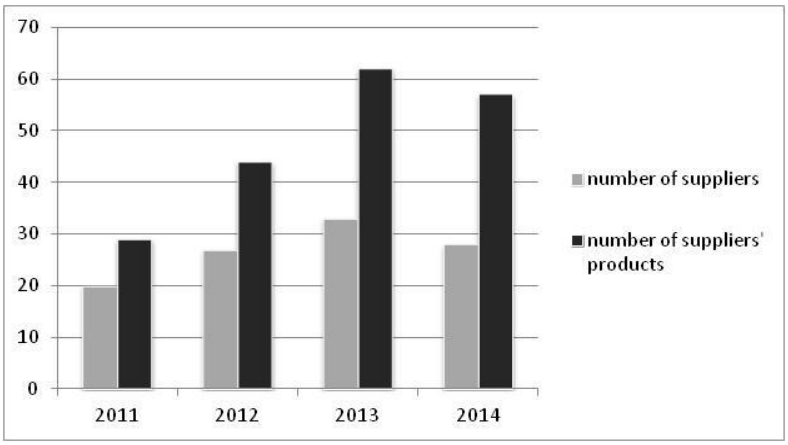


Figure 1 – Comparison of the number of suppliers and their products in period 2011-2014 [3]

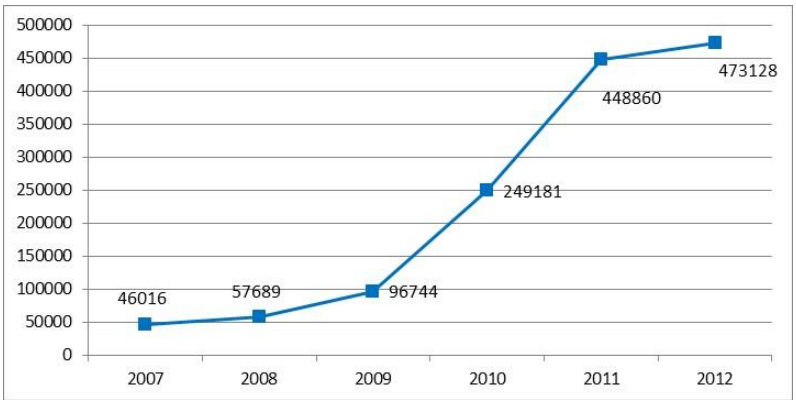


Figure 2 – Number of executed changes of electricity suppliers [10]

Sometimes it is hard to find out the exact price as it is given by more factors such as consumption, fixed fees or taxes. Generally the price can be divided into two components. The first one is the controlled charge for services related to electricity transport from the generator to the final customer. This charge is annually given by Energy Regulatory Office (ERU) [7]. It covers:

- monthly lease for the circuit breaker,
- price per megawatt hour (MWh) in high tariff (HT),
- price per megawatt hour in low tariff (LT),
- price per system services,
- price for the support of the renewable energy purchase,
- charges for the electricity market operator,
- value added tax (with electricity tax 28.30 CZK per 1 MWh).

The second part of the total price is given by the electricity supplier. It covers:

- fixed monthly fee for the selected product,
- price per megawatt hour (MWh) in high tariff (HT),
- price per megawatt hour in low tariff (LT),
- electricity ecological tax (34.24 CZK per 1 MWh including VAT),

The final price is increased by VAT that was 20% till 2012 and 21% from 2013.

If we calculate the average prices of all products offered by the suppliers in each year (Figure 3, left) we can see that the trend till 2012 was increasing but from 2013 it has been falling. One of the arguments for this change can be the changing behavior of households that have started to find the best price. Another one can be the increase of the VAT when suppliers have been afraid of the decrease of customers because of the higher electricity costs and the knowledge of rising distributors prices that raised mainly in 2013 compared to 2012 (Figure 3, right). Various tariff rates are offered to households but their selection is limited as there are limitations for the tariff usage. For this article the tariff rate D25d has been chosen as we use a bigger house as an example (it is assigned when the household uses electric water heater) taken from the work [9]. In this case they can use low

and high tariff. Except of this rate the other standard ones for households are D01 (lower consumption with high tariff only) or D02 (middle consumption with high tariff only). The prices of distributors for each distribution area and the average prices of all suppliers (D25d tariff rate) for the selected period 2011-2014 are in Table 1.

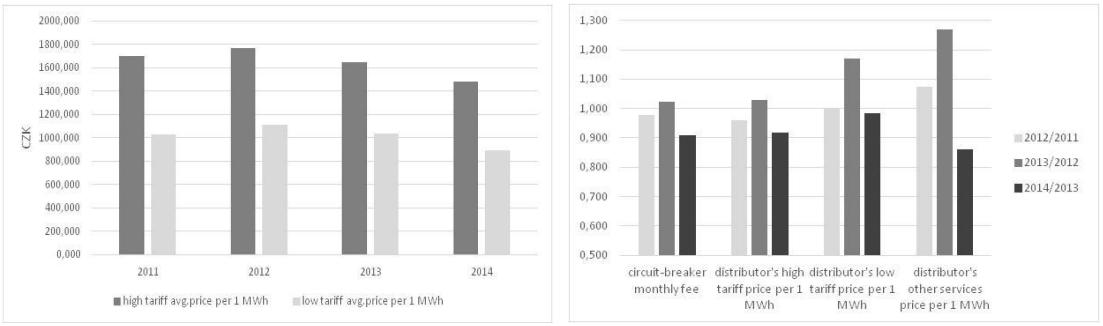


Figure 3 – Comparison of the average prices of high and low tariffs of all suppliers and distribution areas in CZK (left) and the ratio of distributors prices (right) (own calculations, data from [3])

year	distrib. region	Suppliers mothly fee avg.	high tariff avg.price per 1 MWh	low tariff avg.price per 1 MWh	circuit-breaker monthly fee	distributor's high tariff price per 1 MWh	distributor's low tariff price per 1 MWh	distributor's other servi-ces price per 1 MWh
2011	**E.ON**	44.241	1722.172	1022.966	105	1846.39	27.63	530.15
	PRE	48.069	1666.310	1046.759	98	1582.83	19.90	
	CEZ	42.483	1716.241	1008.897	120	1978.50	32.85	
2012	**E.ON**	40.295	1794.136	1098.864	98	1667.65	27.63	569.97
	PRE	47.886	1730.114	1139.091	98	1553.79	19.90	
	CEZ	40.295	1786.886	1087.886	120	1972.84	32.89	
2013	**E.ON**	43.468	1674.666	1021.788	98	1697.42	30.08	722.75
	PRE	50.983	1598.650	1059.587	105	1650.04	25.49	
	CEZ	42.532	1659.016	1017.321	120	1991.98	37.36	
2014	**E.ON**	43.616	1508.339	893.272	90	1592.04	30.59	621.8
	PRE	50.500	1462.345	896.622	98	1563.66	24.45	
	CEZ	44.675	1484.310	886.743	105	1731.93	36.38	

Table 1 – Comparison of prices in CZK (own calculations, data from [3])

3 Simulation Models

Simulation is a technique for imitation of some real situations, processes or activities that already exist in reality or that are in preparation – just to create a computer model [1]. The model can be created in various programs according to the reality and with respect to the problem that should be solved. In fact simulation itself does not provide the solution of the problem but it shows what can happen. It is used to study the system and see how it works, to find where the problems come from, to compare more model variants and select the most suitable one, to show the eventual real effects of alternative conditions and courses of action, etc. [2]. Simulation models can be applied in situation when some variables of the model are uncertain. One powerful tool is known as Monte Carlo simulation that uses more random experiments to find out the possible outcomes. This is typical situation for various decision-making processes in finance [8], banking [6], mobile phone tariffs selection [5], etc. Monte Carlo simulation in connection with the electricity market is usually applied for the random generation of the whole demand for the distributed units [4]. The process of simulation involves a lot of experiments when random number generator and the transformation of the random numbers into random variables from the selected distribution must be used. The spreadsheet add-in package Crystal Ball is designed specifically for Monte Carlo simulation in MS Excel and it has been used also in this article.

4 Results and Discussion

The simulation model can be created according to the anticipated electricity consumption and with respect to the rules of price calculation. We have taken the real data from [9] where the ranges for the electricity consumption in each month were set (at about 900 kWh per month on average). Our simulated consumption has been generated for each month from the normal distribution with 20% of the average taken as the standard deviation. In all

Monte Carlo simulations 1000 experiments have been tried to randomly select consumption for each month and afterwards the annual costs are calculated. From the real data it has been observed that the high tariff is used in 45% from the whole consumption. So the formula for the annual cost calculation for each supplier's product is following:

$$\boldsymbol{COST_{ij}} = \\ = (\mathbf{1} + \boldsymbol{VAT}) \cdot \left[\mathbf{12} \cdot \left(\boldsymbol{mf_{ij}} + \boldsymbol{mf_j}\right) + \mathbf{0.45} \cdot \boldsymbol{gc} \cdot \left(\boldsymbol{ph_{ij}} + \boldsymbol{ph_j}\right) + \mathbf{0.55} \cdot \boldsymbol{gc} \cdot \left(\boldsymbol{pl_{ij}} + \boldsymbol{pl_j}\right) + \boldsymbol{gc} \cdot (\boldsymbol{os} + \boldsymbol{t})\right]$$

where
i … product, i =1, …, n (n = 29 in 2011, n = 44 in 2013, n = 62 in 2013, n = 57 in 2014),
j … distributor, j = 1, …, 3,
VAT … value added tax (VAT = 0.2 in 2011 and 2012, VAT = 0.21 in 2013 and 2014),
mf … fix monthly fee,
gc … yearly generated consumption in MWh,
ph … price in high tariff per 1 MWh,
pl … price in low tariff per 1 MWh,
os … price for other services per 1 MWh,
t … electricity tax per 1 MWh (t = 28.3 CZK for all years).

We have found out that the difference among the distributors regions exist as the final annual price for the same product of one supplier is different in relation to the distribution region in all years (example – Figure 4, right). In 2011 the cheapest product is offered by the United Energy Trading company in PRE region. The most expensive product is the "E.ON elektrina trend aku prosinec" which is more expensive of about 3000 CZK (10%) on the average (Figure 4, left). The PRE distribution region is always the cheapest one for all products in all years (see example – Figure 4, right). The lowest price has been offered by the Amper Market company in 2012, by the CEZ in 2013, the highest by the Central Energy company in 2012 and by the Global Energy company in 2013 (Figure 5).

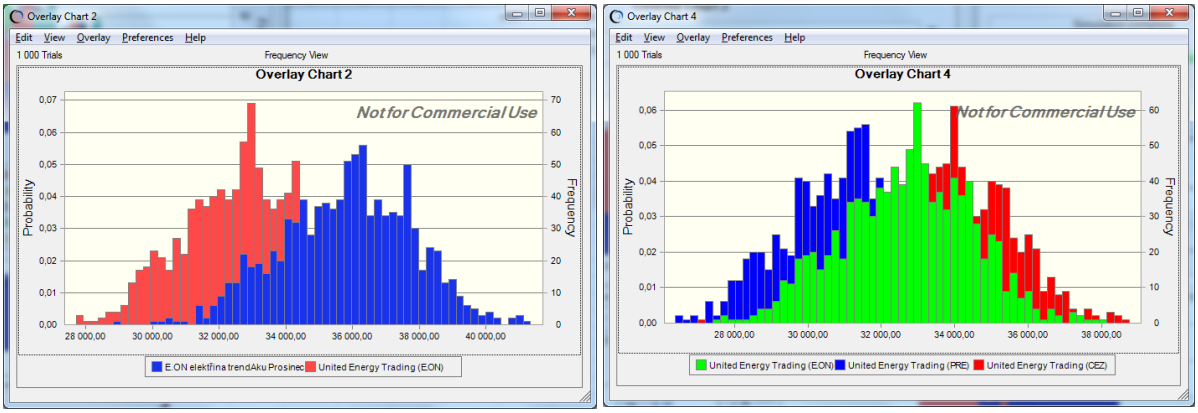


Figure 4 – Best and worse suppliers products in 2011 (left – E.ON region) and the best product in different distributors regions in 2011 (right)

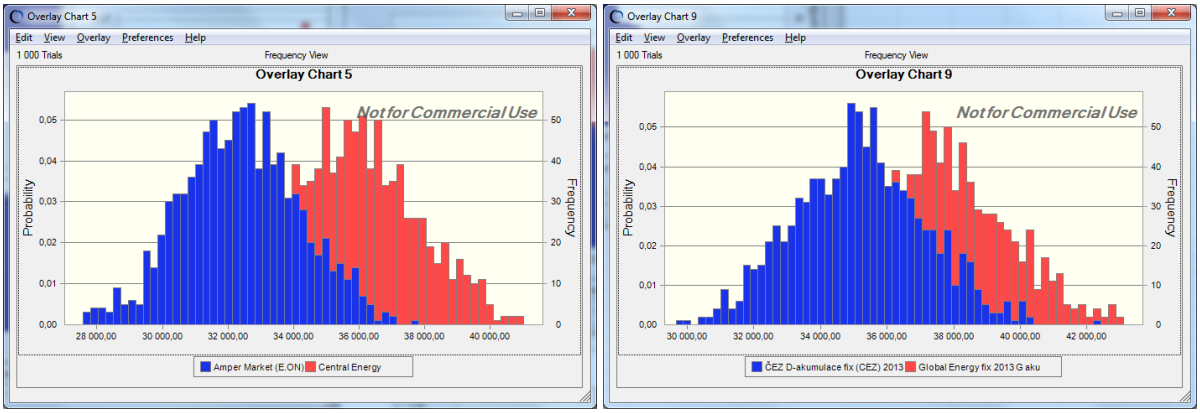


Figure 5 – Best and worse products in 2012 (left) and 2013 (right)

When we compare final annual costs of all suppliers' products in one year (example of 2014 – Figure 6) it is obvious that the differences in the average annual costs for the given situation exist. Some of the suppliers (Bohemia Energy, ELIMON, Global Energy) set their prices higher than others and so the average annual costs are about 10-16% higher than in case of the cheapest products. On the other hand the 90% certainty bounds (when the consumption is normally distributed) overlap for all products. Table 2 describes the cheapest five products (or the supplier is mentioned when just one product is offered) in the selected years. Only Amper Market and Europe Easy Energy suppliers figure in „Top 5" in more than one year. On the basis of the prices of suppliers and distributors (Figure 3) we see the same trend in the lowest and highest prices for the simulated consumption (Figure 7, left). The comparison of the selected products that have been offered during the whole period 2011-2014 shows (Figure 7, right) that the trend is nearly the same for all of them but the Amper Market and Europe Easy Energy are still the cheapest. The peak in 2013 is cause mainly by the higher distributors prices and also by the higher VAT. But in 2014 the range of the annual cost is much more smaller than in 2011. This fact indicates that the electricity market is going to be stable and the final electricity consumption annual costs are directed into the similar values.

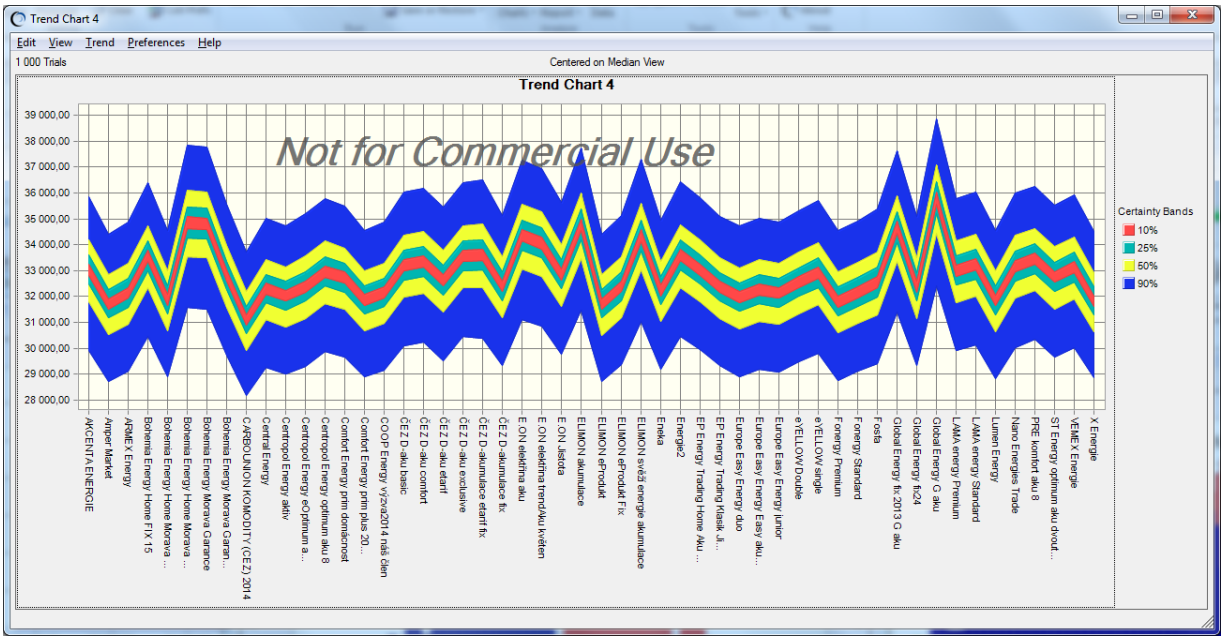


Figure 6 – All products and the final prices ranges (2014)

Order	2011	2012	2013	2014
1	United Energy Trading	Amper Market	CEZ D-aku fix	CARBOUNION KOMODITY
2	Lumen Energy	Europe Easy Energy	ELIMON eProdukt Fix	ELIMON eProdukt
3	Amper Market	Comfort Energy	ČM energetika	Amper Market
4	X Energie	EP Energy Trading	Europe Easy Energy	Lumen Energy home aku
5	Europe Easy Energy	Fosfa	CEZ D-aku etarif fix	Fonergy Premium

Table 2 – Final results – the cheapest 5 products in the period 2011-2014 (PRE region, tariff D25d)

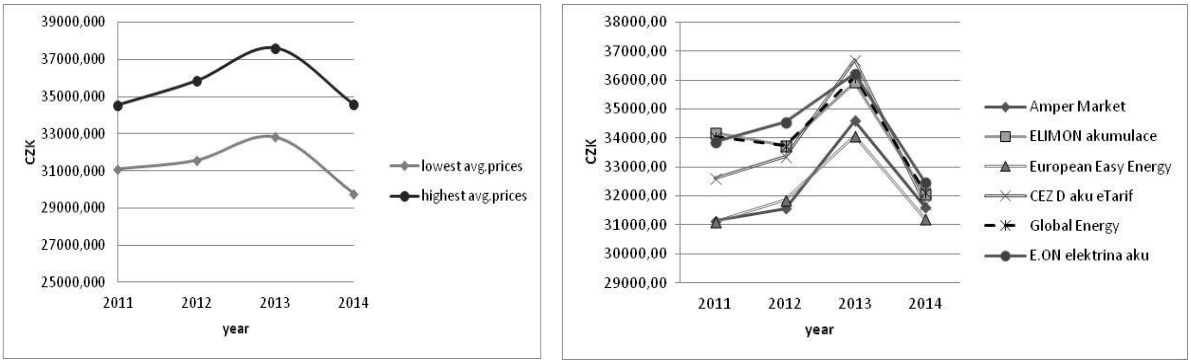


Figure 7 – Lowest prices in PRE region for each year (for products see Table 2) – left; comparison of the final annual price of the selected suppliers' products (PRE region) – right

5 Conclusions

The situation on the electricity retail market in the Czech Republic is not transparent because of the number of suppliers and its products. This number has increased till 2013, in 2014 it is falling down slowly. Each household can choose its electricity supplier on the basis of the conditions offered but the final costs are influenced also by the distribution region. For all products the cheapest distributor is PRE. During the period 2011-2014 we see different trend of suppliers' and distributors' prices as the suppliers reached their peak in 2012 but the distributors in 2013. Regarding the fact that the formula of the annual cost calculation of the electricity consumption contains a lot of factors it is hard to choose the cheapest product. Monte Carlo simulation of the electricity consumption might help in this situation as it considers uncertainty in the consumption. The results have shown that the situation on the retail market is starting to be stable and the prices are closer to each other than in 2011. On the other hand it is impossible to recommend definitely the strategy of the cheapest product selection in view of the fact that some "winners" have left the market next year. But each household should monitor the situation on the market as the exchange of product can save more than 10% of annual electricity costs.

Acknowledgements

The research project was supported by Grant No. 13-07350S of the Grant Agency of the Czech Republic and also by the grant No. F4/11/2013 of the Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Banks, J.: *Handbook of Simulation*. John Willey & Sons, USA, 1998.
[2] Dlouhý, M., Fábry, J., Kuncová, M. and Hladík, T.: *Simulace ekonomických procesů*. 2. ed. Computer Press, Brno, 2011.
[3] Energy Regulatory Office (ERU): *Price calculator* [online], available at: http://kalkulator.eru.cz/VstupniUdaje.aspx [cit. 2014-05-2]
[4] Hegazy, Y.G., Salama, M.M.A. and Chickhani, A.Y.: Adequacy assessment of distributed generation systems using Monte Carlo Simulation. *Power Systems - IEEE Transactions* 18/1 (2003), 48-52.
[5] Kuncová, M.: Practical Application of Monte Carlo Simulation in MS Excel and its Add-ons – The Optimal Mobile Phone Tariffs for Various Types of Consumers in the Czech Republic. In: *Proceedings of Mathematical Methods in Economics*, Pilsen, 2006, 323-332.
[6] Kuncová, M. and Lízalová, L.: Monte Carlo simulation – the bank account selection in the Czech Republic according to the bank charges. In: *Proceedings of the International Conference on Modeling and Applied Simulation*. Wien, 2012, 153-161.
[7] *National Report of the Energy Regulatory Office on the Electricity and Gas Industries in the Czech Republic for 2012* [online], available at: http://www.eru.cz/documents/10540/488714/NR_2012.pdf/1f82372e-bfd8-43e7-8ce7-086bebaf7481 [cit. 2014-04-28]
[8] Razgaitis, R.: *Dealmaking Using Real Options and Monte Carlo Analysis*. John Willey & Sons, New Jersey, 2003.
[9] Rejmanová, V.: *Volba dodavatele elektřiny pro domácnost s využitím metod operačního výzkumu*. VŠE Prague, diploma thesis, 2013.
[10] *Year Report on the Electricity and Gas Market in the Czech Republic for 2012* [online], available at: http://www.ote-cr.cz/o-spolecnosti/soubory-vyrocni-zprava-ote/rocni-zprava-2012.pdf [cit. 2014-04-30]

Search for the optimal strategy to spread a viral video: An agent-based model optimized with genetic algorithms

Michal Kvasnička[1]

Abstract. Agent-based computational papers on viral marketing have been so far focused on the study of the word-of-mouth knowledge diffusion that merges the decisions to adopt a product and to share information about it. This approach is not suitable for the analysis of the viral video sharing because it is shared with no regard whether the sender has adopted the advertised product or not. This paper presents a more realistic model of viral video diffusion in which every agent that viewed the video shares it once with a random subset of her neighbors. The optimal seeding strategy is then searched with genetic algorithms. The seeding strategy found by the genetic algorithm includes into the initial seed the agents with most connections and lowest clustering ratios; some agents are also selected randomly. However, this complex seeding strategy does not perform significantly better than a simple strategy of selecting agents with many connections.

Keywords: viral video, viral marketing, social network, agent-based model, genetic algorithm.

JEL classification: C61, C63, D85, M31
AMS classification: 68U20, 37N40, 05C82, 90B15

1 Introduction

All previous agent-based models of viral marketing merge the individuals' decisions to adopt a new product and to share information about it. This approach is suitable to model the traditional word-of-mouth diffusion of information but it does not seem to capture well the diffusion of viral videos over a social network. It is because the decisions to adopt a product and to share a video advertising it are separated: an individual can share the video because she regards it as entertaining, interesting, or alarming without adopting the product. This paper provides an alternative model of viral content diffusion that is better suited to capture this feature of viral video sharing. The goal of this paper is to explore by means of an agent-based computational simulation what constitutes an optimal seeding strategy in viral video marketing, i.e. which and how many agents the marketer should initially "infect" with the video when she has to pay for accessing them. The optimal seeding strategy is searched by genetic algorithms in the way proposed by Stonedahl, Rand, and Wilensky [12] for the ordinary "word-of-mouth" viral marketing.

2 Review of the existing literature on agent-based viral marketing

There is a wide agent-based computational literature on viral marketing. (For an introduction into the agent-based computational modeling, ACE, see e.g. [13]; for its typical application, see e.g. [8].) The ACE models of viral marketing consist of three parts: an explicit description of the used social network, the activation mechanism (i.e how an agent gets ready to share the video, i.e. get "infected"), and the seeding strategy (i.e. which and how many agents get initially infected by the marketer).

The researchers use both various artificial and empirical networks. Among the artificial networks, the most used ones are random network [12, 14], lattice and its modifications [1, 6, 12], ring [9], small-world [12], power network and its modifications [3, 9, 12, 14], mix of ring and power network [9], and a network consisting of small complete sub-networks connected by scarce inter-cluster connections [5, 14]. Several empirical networks have been used too: a network consisting of a segment of Twitter users [12], a network of users of the Korean social network Cyworld [4], and a network of coauthors of scientific papers [2, 7].

[1]Masaryk University, Faculty of Economics, Department of Economics, Lipová 41a, Brno, 602 00, michal.kvasnicka@econ.muni.cz

Great majority of researchers uses one of two following activation mechanisms: either a variant of the threshold model or a variant of the epidemiological SIR model. In the threshold model, an agent gets infected when at least x her neighbors are already infected, where x can be either an absolute number or a relative share of her neighbors. In the SIR model, an agent gets infected with some probability every time she interacts with any of her infected neighbors; thus the probability she gets infected rises to unity over time. The pure threshold activation is used e.g. in [1, 2, 7, 12, 14], the pure SI action is used e.g. in [2, 7, 14]; the combination of the threshold and personal preference of the agents is used in [3, 6], a weighted mix between SI and threshold is used in [9], and a combination of an external contamination (an advertisement) and SI (the word-of-mouth) is used in [5]. In all these approaches, the knowledge is shared only by the adopters of the product, i.e. being infected means that an agent has bought and consumes the product. The intuition of the diffusion is such that an agent buys a product because she can see some of her neighbors consuming it.

The usual seeding strategy is such that one or more agents are randomly infected at the beginning of the simulation. Other randomly chosen agents can get infected later in the simulation run to mimic the impact of an advertisement, as in [5]. Few papers try to find the optimal seeding strategy and the optimal seed size. The optimization requires perfect knowledge of the network and an immense computational power. The formalization of the problem and an approximate algorithm to solve it for a given seed size is provided by [7]. However, it is more realistic to assume only local information about the agents such as number of their connections (degree), their clustering ratio, etc. Stonedahl, Rand, and Wilensky [12] use genetic algorithms to solve this "local viral marketing problem". Their strategies consist of the seed size and weights placed on agents' desirable properties. They infect the selected number of agents with the highest weighted sum of measures of these properties. There is a great variability in their optimal strategies but the most important factors to select an agent into the seed are her degree (i.e. the number of her connections) and her clustering ratio. The optimal seed size is the lower the higher is the Gini cofficient of agents' degree (i.e. the more unequally the degree is distributed). However, their optimal strategies do not perform significantly better than a simple seeding strategy that selects the agents with the highest degree.

3 Model

3.1 Activation mechanism

I call every agent that has viewed the video *infected* with no regard whether she has adapted the advertised product or not. An infected agent's decision whether and with whom to share a video depends on three determinants: how much the agent is used to share videos in general, how much she finds the video appealing, and how much she believes her neighbors would like to see it. The decision to view a video shared by another person depends on one's personality and the relationship she has with the sender. In this paper, these two complex decisions are modeled as a simple probabilistic act: an infected person shares the video with each of her neighbors with some probability. If she shares it with a person, the person gets infected. Each infected person shares the video only once. The precise mechanism of the activation is following: At the initialization, every agent i draws a probability p_i that she shares the video; p_i is drawn from the continuous uniform distribution $U(0, v)$ where v is the maximal *virality* of the video. Then she creates a list l_i of agents to share the video with: she adds each of her neighbors to the list l_i with probability p_i. When agent i is first infected at time t, she shares the video with all agents in her list l_i at time $t + 1$, and each of these agents get instantly infected. Agent i shares the video with no one after time $t + 1$.

3.2 Network structure

An agent's neighborhood is defined by a social network in which the agent is located. Actual internet social networks seem to consist of two kinds of relations between the agents: friendship and following. Friendship is a symmetric relationship between two agents that can share stuff, e.g. videos, with each other. The stylized facts are that friends are highly clustered (i.e. one's two friends are likely to be friends together), the mean length of path between any two agents is quite short (about seven), and the number of one's friends has a power distribution, i.e. many people have few friends while few people have many friends. Following is an asymmetric relationship between a followed person and a follower, e.g. between a celebrity and her fan. The followed person can share stuff (e.g. videos) with the follower but not vice versa. The stylized fact is that the number of one's followers has a power distribution, i.e. most agents have very few followers while very few agents have very many followers. For the sake of simplicity, I assume that no person can be at the same time one's friend and follower.

The model uses an artificial social network that tries to replicate these stylized facts. Since there is no widely accepted algorithm to generate such a network, the two most widely used types of network, small world and power network, are mixed together to get a network with the suitable properties. The small world network is highly clustered and its average path length is small but its degree distribution is rather symmetric. The power network has the short average path length property and its nodes' degrees have a power distribution but its nodes are not clustered.

Each model network consists of 1 000 agents and its parameters are selected to resemble the properties of the empirical networks. Each network is created in two steps. First, a small world network is created by algorithm described by [15]; the code has been adapted from [17]. This network consists of symmetric (i.e. undirected) links that represent friendship. The agents are arranged into a circle. Each agent initially has 10 friends, 5 agents to the left of her and 5 agents to the right of her. Each link representing friendship is then rewired with probability 10 %, which creates the initial small world network. Second, a power network of directed links representing the following relationship is created over the friends' network. The algorithm has been adapted from [18]: one follower connection is added at a time, each agent is selected randomly as the follower with an equal probability and one other agent is selected randomly as the followed one with the probability proportional to each agent's number of friends and followers. 2 000 followers' links (i.e. 2 links per agent) are added. The mean clustering ratio of the networks is about 35 %. The distribution of the number of friends is symmetric, while the distribution of the number of followers has roughly a power distribution. The set of agent i's neighbors used in the activation mechanism described above is the union of the set of her friends and the set of her followers.

3.3 Marketer's seeding strategy and profit

The marketer's seeding strategy used in this paper is based on the approach developed by [12]. The agents are included into the seed because they have some desirable properties. Since there are multiple desirable properties, they are weighted. A seeding strategy (S, w) thus consists of two parts: the seed size S, i.e. how many agents the marketer initially infects, and weights w_j placed on measures of some desirable properties f_j that determine which agents are selected into the seed. An index $\sum w_j f_j$ is calculated for each agent and S agents with the highest value of the index are included into the seed. I use six measures f_j: 1) f_1 = the number of agent's friends divided by the maximal number of friends in the population, 2) f_2 = the number of agent's followers divided by the maximal number of followers in the population, 3) f_3 = the number of agent's friends and followers divided by the maximal number of friends and followers in the population, 4) f_4 = 1− (agent's clustering ratio / the maximal clustering ratio in the population), 5) $f_5 = f_3 \cdot f_4$, and 6) f_6 is a random number drawn from $U(0, 1)$. The first three f_js are various measures of degree – the more connections an agent has, the better she can share the video. The f_4 is the agent's clustering ratio – the less likely an agent's neighbors are to be neighbors themselves, the better the video can spread through the population. The f_5 generalizes the notion of clustering and degree – the maximal degree and the minimal clustering are beneficial at the same time. The f_6 allows including agents into the seed randomly, which may be beneficial e.g. when the agents with the highest degree are connected together.

In contrast to the previous agent-based models of viral marketing, I explicitly assume that an agent's decisions to share a viral video and to adopt the product advertised by the video are independent. The expected revenue from the viral marketing campaign is then equal to $\rho\sigma N$, where N is the number of agents infected during the campaign, ρ is the probability that an infected agent adopts the product because of the campaign, and σ is the profit from one adopter. The cost of the campaign is $\gamma S + F$ where γ is a cost of seeding one agents, S is the seed size, and F is a fixed cost, e.g. the cost of creating the video. The marketer's problem is then to select the seeding strategy (S^*, w^*) that maximizes her expected profit from the campaign, $\Pi = \rho\sigma N - \gamma S - F$. The seeding strategy (S^*, w^*) maximizing Π also maximizes $\pi = N - cS$ where $c = \gamma/(\rho\sigma)$ is the relative cost of seeding one agent. (I do not explicitly address the problem of setting the optimal budget for the video creation. Thus F, ρ, and v are treated as constants.)

3.4 Simulation, implementation, and optimization

Each simulation consists of two parts: the initialization and the run. In the initialization, the agents are created and connected within the model network. Each agent i is assigned the probability p_i that she shares the video with her friends and followers. She then creates the list l_i of the agents which she shares the video with if infected. At the end of the initialization, the initial agents are infected. Each simulation run proceeds in discrete steps. In each step, each agent i infected in the previous step infects the agents in l_i and the total number of infected agents and the marketer's relative profit π are calculated. The run ends when there are no infected agents that have not yet shared

the video. The model was simulated for 1 000 agents, the activation and network described in sections 3.1 and 3.2, maximal virality $v = 20$ %, and relative seeding cost $c = 10$. The model was implemented in NetLogo 5.0.5 [16]. The web interface of the model is available at http://www.econ.muni.cz/~qasar/english/models.html.

Since the parameter space of the problem is huge, the optimal seeding strategy (S^*, w^*) was searched by a genetic algorithm (GA). The search was performed in BehaviorSearch 1.0 [11]. The seed size S was searched on domain of $1, 2, \ldots, 50$, each weight w_j on domain $0, 0.01, \ldots, 1$. As in [12], all variables were encoded as Gray binary chromosomes. The initial population of strategies consisted of 50 randomly chosen strategies. The standard generational evolution steps (one-point crossover rate 0.7, mutation rate 0.03, and tournament selection with tournament size 3) were performed on the population of the strategies for 200 generations. The fitness of each individual strategy was evaluated by the mean relative profit π of 20 independent replications of the simulation. The best strategy was selected based on the recheck of 50 independent simulations. The individual weights were re-scaled to sum to unity. The whole search for the optimal parameters was repeated 30 times. In total, the search simulated roughly 6 million individual models, which took about 480 CPU-days.

To evaluate the relative performance of the GA optimal strategies, each optimal strategy and selected simple strategies were simulated on 100 additional instances of the model network. The simple strategies involved including $S = 1, \ldots, 20$ agents with the highest number of friends, followers, or the sum of friends and followers into the initial seed. The resulting data were analyzed in R 3.1.0 [10].

4 Results

The GA optimal strategies found for our type of activation, social network, and profit specification resemble the results of [12]. The variability of the optimal weights is huge (see Figure 1a), while the expected profit is very similar for all optimal strategies (see Figure 2). This indicates that there is a partial substitutability between the measures of the desirable properties f_j and perhaps also that there is a plateau in the profit function. The first conjecture is supported by the fact that the variability of the sum of weights of all degree measures and the weights of the two clustering measures is much lower (see Figure 1b). In general, the optimal strategies place a great weight on the measures of degree and a smaller weight on the measures of clustering; surprising is the relatively high weight put on the random selection, which is in a stark contrast to the findings in [12]. This may further support the conjecture about the plateau in the profit function. The optimal seed sizes are relatively small: roughly equal to 1 % of the whole population of agents (see Figure 1c). This is probably caused by the relatively high relative cost c.

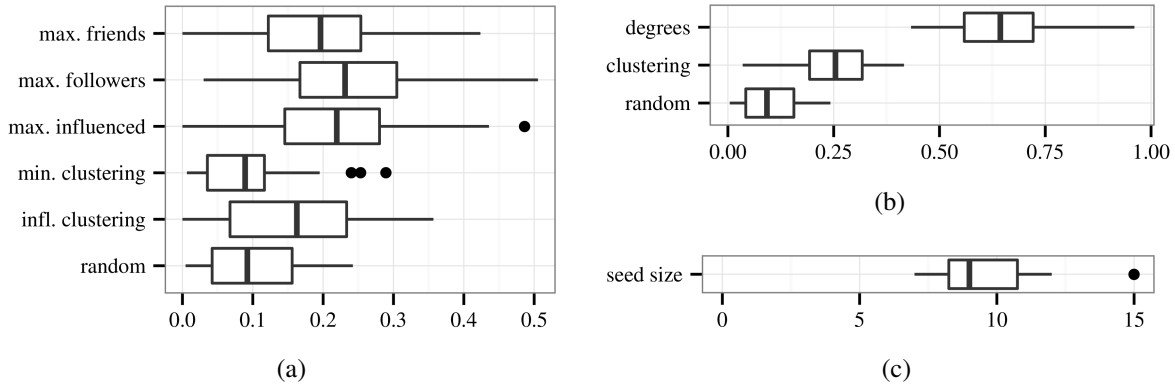


Figure 1 Distribution of the GA optimal seeding strategies found by the genetic algorithm. (a) Individual weights w_j. (b) Groups of weights: degrees are the sum of the weight of the maximal number of friends, followers, and their sum; clustering is the sum of the weights of the minimal clustering and product of maximal number of friends and followers and minimal clustering. (c) Seed sizes.

The expected relative profits π of all GA strategies are similar. However, the variance of the profits is huge (i.e. the viral video marketing campaign is risky); the same holds true for the simple seeding strategies (see Figure 2). Moreover, the GA optimal seeding strategies do not perform much better than the simple strategies with the appropriate seed size. This can be more clearly seen in Figure 3: even the seeding strategy with weights taken from the GA optimal strategy that performed the best on the sample of 100 networks does not perform better than the simple strategy of seeding agents with most followers or most friends and followers.

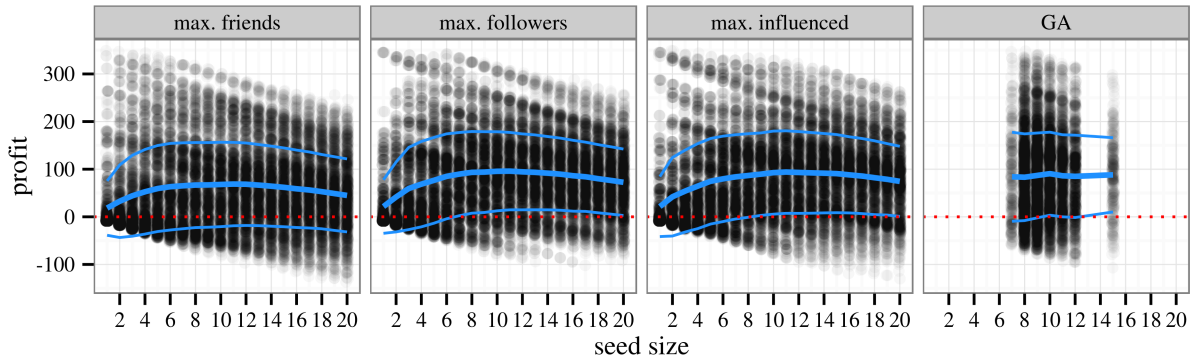


Figure 2 Comparison of performance of the GA optimal strategies and simple types of strategies (selecting agents with maximal number of friends, followers, or their sum). Individual semitransparent dots denote the relative profits π in individual simulations. The thick solid line denotes the average profit for each seeding type and seed size. The thin solid lines denote the average profit plus minus one standard deviation. The dotted line denotes the zero profit.

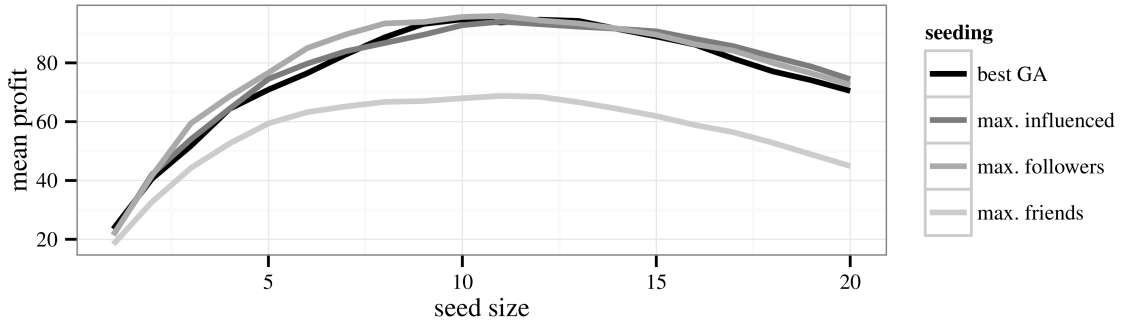


Figure 3 Comparison of expected relative profits π of the simple strategies and the best performing GA optimal strategy (GA strategy seed sizes were changed to be comparable with the simple strategies).

5 Conclusion

Our findings provide some external verification for the results of [12]. Even though I used a different type of a social network, activation mechanism, and profit specification, the optimal seeding strategy found by the genetic algorithm is similar to that of [12]: it puts a great weight on measures of degree and to a smaller extent on measures of clustering. Also similar to [12], the optimal GA strategy performed no better than a simple strategy of including agent with many connections into the initial seed. However, these results must be seen as preliminary since they were achieved only for one particular (rather low) virality of the video and one particular (rather high) relative cost of seeding. Their generalizability should be a subject of further research.

Acknowledgements

This paper has been created as a part of project of specific research no. MUNI/A/0781/2013 at the Masaryk University.

Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the program "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005), is greatly appreciated.

References

[1] Brudermann, T., Fenzl, T., 2010. Agent-based Modelling: A new approach in viral marketing research. In: *Advances in Advertising Research* (Terlutter, R., Diehl, S., Okazaki, S., eds.), vol. 1. Springer Gabler, 2010, 397–412.

[2] Cointet, J. P., and Roth, C.: How realistic should knowledge diffusion models be?. *Journal of Artificial Societies & Social Simulation* **10**(3), 2007.

[3] Delre, S. A., Jager, W., Bijmolt, T. H., and Janssen, M. A.: Will it spread or not? The effects of social influences and network topology on innovation diffusion. *Journal of Product Innovation Management* **27**(2), 2010, 267–282.

[4] Goldenberg, J., Han, S., Lehmann, D. R., and Hong, J. W.: The role of hubs in the adoption process. *Journal of Marketing* **73**(2), 2009, 1–13.

[5] Goldenberg, J., Libai, B., and Muller, E.: Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters* **12**(3), 2001, 211–223.

[6] Goldenberg, J., Libai, B., Solomon, S., Jan, N., and Stauffer, D.: Marketing percolation. *Physica A: Statistical Mechanics and its Applications* 284(1), 2000, 335–347.

[7] Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2003, 137–146.

[8] Krčál, O.: An agent-based model of price flexing by chain store retailers. In: *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012* (Ramík, J., Stavárek, D., eds.), Silesian University, 2012, 461–466.

[9] Kurahashi, S., and Saito, M.: Word-of-mouth effects on social networks. In: *Knowledge-Based and Intelligent Information and Engineering Systems* (König, A., Dengel, A., Hinkelmann, K., Kise, K., Howlett, R. J., and Jain, L. C., eds.), part 3, Springer, 2011, 356–365.

[10] R Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. http://www.R-project.org/.

[11] Stonedahl, F. BehaviorSearch, computer software, 2010. http://www.behaviorsearch.org/.

[12] Stonedahl, F., Rand, W., and Wilensky, U.: Evolving viral marketing strategies. In: *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, ACM, 2010, 1195–1202.

[13] Tesfatsion, L., 2006. Agent-based computational economics: a constructive approach to economic theory. In: *Handbook of Computational Economics* (Tesfatsion, J., and Judd, K. L., eds.), vol. 2, 2006, Elsevier, 831–880.

[14] Watts, D. J., and Dodds, P. S.: Influentials, networks, and public opinion formation. *Journal of consumer research* **34**(4), 2007, 441–458.

[15] Watts, D. J., and Strogatz, S. H.: Collective dynamics of 'small-world' networks. *Nature* 393, 1998, 400–442.

[16] Wilensky, U.: NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999. http://ccl.northwestern.edu/netlogo/.

[17] Wilensky, U.: NetLogo Small Worlds model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 2005. http://ccl.northwestern.edu/netlogo/models/SmallWorlds.

[18] Wilhite, A.: Economic activity on fixed networks. In: *Handbook of Computational Economics* (Tesfatsion, J., and Judd, K. L., eds.), vol. 2. Elsevier, 2006, 1013–1045.

Price of Fairness in Public Service System Design

Marek Kvet[1], Jaroslav Janáček [2]

Abstract. In the host situations, a public service system is designed so that the total disutility, like social costs, is minimized. The social costs are often proportional to the total distance travelled by all users to the nearest service center. This approach often causes such situation that the total social cost is minimal, but disutility of the worst situated user is extremely high, what is denoted as unfair system design. A fair approach to the public service system design consists in the process, when the disutility of the worst situated users is minimized first and then disutility of better located users is optimized subject to the disutility of the worst situated users does not worsen, what is called lexicographical minimization. In this contribution we present an approximate approach based on radial formulation of the problem with homogenous system of radii given by dividing points. This approach will be used for both system optimal public service system design and the fair one with specific ways of dividing point deployment. In the computational study, we present the solution of several real instances and evaluate the price of fairness connected with two different designs.

Keywords: Price of fairness, radial approach, lexicographical optimization

JEL Classification: C61
AMS Classification: 90C06, 90C10, 90C27

1 Introduction

A logistic or system approach to the public service system design (PSSD) consists in deployment of at most given number of service centers at some places of a finite set of possible locations so that the sum of disutility values perceived by individual system users be minimal [2], [8], [11]. Even if all system users share the costs for the system construction, not all users have the same access to the service. Especially, users from rear populated dwelling places do not consider this solution fair, because their contribution to the system cost are equal to the contribution of the others, but their distance from the nearest service center is much bigger than an average disutility in the service system. To comply with the user's call for equity in the service accessibility various solidarity criteria were developed [1], [12]. The strongest one of the criteria is the lexicographic min-max approach, which minimizes the disutility of the worst situated user, and then the disutility of the second worst situated user is minimized unless the minimal reached disutility of the previously processed users is worsen. This process provides us with the lexicographic optimal solution. It is obvious that the fair solution is paid by a loss of the average user's disutility. To express the loss, a measure called the price of fairness (POF) was introduced [1]. The public service system design is closely connected with the weighted p-median problem solution. We concentrate our effort into such way of the design, which employs a commercial IP-solver and thus enables to avoid a designer's long term development of the optimization technique and the associated software. For this way of design, two approaches are available. The first one describes the PSSD by location-allocation model and the second one uses so-called radial model [3], [5]. As the first approach often fails due to enormous demand on computer memory or computational time, we focus on the radial approach with homogenous system of radii defined by a series of the dividing points [7]. Having defined a series of the dividing points for the system optimal PSSD, we make use of the series and developed the computational process for the fair optimal PSSD. The process is based on the Ogryczak's general approach [12], but we make use an advantage of the radial model. The both system and fair optimal public service system designs can be obtained using the radial approach and compared from the point of the price of fairness and the Hamming distance between associated solutions. The remainder of the paper is organized as follows. The second section is devoted to a concise description of the radial formulation of the weighted p-median problem based on the dividing points. The third section contains the radial formulation used together with Ogryczak's approach. The notion of the price of fairness and the Hamming distance are ex-

[1] Ing. Marek Kvet, PhD., University of Žilina, University Science Park, Univerzitná 8215/1, 010 26 Žilina, Slovakia, e-mail: marek.kvet@uniza.sk
[2] Prof. RNDr. Jaroslav Janáček, CSc., University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, e-mail: jaroslav.janacek@fri.uniza.sk

plained in the fourth section. The fifth and sixth sections are devoted to the computational study performed with the eight regions of Slovakia accompanied by the comparison of the system and fair approaches.

2 Radial approach to the PSSD with system criterion

The public service system design with minimal disutility is a task of location of at most p service centers so that the sum of individual disutility values of each user coming only from the nearest located service center is minimal. To describe the problem, we denote I a set of possible service center locations and J the set of possible users' locations. The symbol b_j denotes the number of users located at the location j. The disutility for a user at the location j following from the possible center location i is denoted as d_{ij}. Under assumption that the user's disutility originates only from the nearest service center, we can state the problem as follows:

The problem (1) - (5) is also known as the weighted p-median problem, which is broadly discussed in [3], [5], [6] from the viewpoint of solving techniques suggested for fast solving of the huge instances. Within this paper, we focus on so called radial formulation used in the above papers and we explore the approaches based on the set of dividing points.

The strategic decision in the problem (1) - (5) concerns location of centers at possible center locations from the set I. To model this decision at particular location, we introduce a zero-one variable $y_i \in \{0, 1\}$, which takes the value of 1, if a center should be located at the location i, and it takes the value of 0 otherwise. To obtain an upper or a lower bound of the original objective function, the range $[d_0, d_m]$ of all possible disutilities $d_0 < d_1 < \ldots < d_m$ from the matrix $\{d_{ij}\}$ is partitioned into $v+1$ zones. The zones are separated by a finite ascending sequence of so called *dividing points* $D_1, D_2, \ldots, D_v$ chosen from the sequence $d_0 < d_1 < \ldots < d_m$, where $0 = d_0 = D_0 < D_1$ and also $D_r < D_{v+1} = d_m$. The zone s corresponds with the interval $(D_s, D_{s+1}]$. The length of the s-th interval is denoted by e_s for $s = 0, \ldots, v$. In addition, auxiliary zero-one variables x_{js} for $s = 0, \ldots, v$ are introduced. The variable x_{js} takes the value of 1, if the disutility of the user at $j \in J$ from the nearest located center is greater than D_s and it takes the value of 0 otherwise. Then the expression $e_0x_{j0} + e_1x_{j1} + e_2x_{j2} + \ldots + e_vx_{jv}$ constitutes an upper approximation of the disutility d_{j*} from user location j to the nearest located service center. If the disutility d_{j*} belongs to the interval $(D_s, D_{s+1}]$, then the value of D_{s+1} is the upper estimation of d_{j*} with the maximal possible deviation e_k. Let us introduce a zero-one constant a_{ij}^s for each triple $[i, j, s] \in I \times J \times \{0, \ldots, v\}$. The constant a_{ij}^s is equal to 1, if the disutility d_{ij} between the user location j and the possible center location i is less or equal to D_s, otherwise a_{ij}^s is equal to 0. Then the radial-type weighted covering model can be formulated according to [5], [6] as follows:

$$\textit{Minimize} \qquad \sum_{j \in J} b_j \sum_{s=0}^{v} e_s x_{js} \tag{1}$$

$$\textit{Subject to}: \quad x_{js} + \sum_{i \in I} a_{ij}^s y_i \geq 1 \quad \textit{for } j \in J,\ s = 0, 1, \ldots, v \tag{2}$$

$$\sum_{i \in I} y_i \leq p \tag{3}$$

$$x_{js} \geq 0 \quad \textit{for } j \in J,\ s = 0, 1, \ldots, v \tag{4}$$

$$y_i \in \{0, 1\} \quad \textit{for } i \in I \tag{5}$$

The objective function (1) gives the upper bound of the sum of original disutility values. The constraints (2) ensure that the variables x_{js} are allowed to take the value of 0, if there is at least one center located in radius D_s from the user location j. The constraint (3) puts a limit p on the number of located facilities.

To obtain a lower bound of the objective function value of the optimal solution of the original problem, several approaches can be employed. We present here the simplest one, which can be introduced as the expression $e_0x_{j1} + e_1x_{j2} + e_2x_{j3} + \ldots + e_{v-1}x_{jv}$.

3 Radial approach to the PSSD with fair criterion

The fair public service system design with fair criterion is a task of location of at most p service centers so that the individual values optimize some fairness scheme. The disutility perceived by an individual user comes only

from the nearest located service center. Under assumption that the user's disutility originates only from the nearest service center, the disutility for a user at the location j can be described as follows:

$$g_j(\mathbf{x}) = \sum_{s=0}^{v} e_s x_{js} \tag{6}$$

Let the range of all disutility values be represented by a finite set of ordered values G_0, G_1, …, G_w, G_{w+1}, so that $G_{max} = G_0 > G_1 > \ldots > G_w > G_{w+1} = G_{min}$. The set of ordered values represents a series of points dividing the range into $w+1$ zones $(G_{u+1}, G_u]$ for $u = 0, 1, \ldots, w$.

The lexicographic min-max problem according to [12] consists in lexicographic minimizing of the vector $[B_0(\mathbf{z}), B_1(\mathbf{z}), \ldots, B_w(\mathbf{z})]$ subject to (2) – (5).

For this case, where $\{G_u\}$ is set of dividing points chosen from set of all general disutility values, $B_u(\mathbf{z})$ is defined by (7).

$$B_u(\mathbf{z}) = \sum_{\substack{j \in J \\ g_j(\mathbf{z}) \in (G_{u+1}, G_u]}} b_j \qquad for\ u = 0, \ldots w \tag{7}$$

The nonlinearity in the expressions (7) was originally excluded by introducing a series of slack non-negative variables h_{tj} for given stage t and for $j \in J$. The variables h_{tj} must be connected to the system of original variables by link-up constraints (8) and obligatory constraints (9).

$$h_{uj} \geq \sum_{i \in I} d_{ij} z_{ij} - G_{u+1} \qquad for\ u = 0, \ldots t,\ \ j \in J \tag{8}$$

$$h_{uj} \geq 0 \quad for\ u = 0, \ldots t,\ \ j \in J \tag{9}$$

The constraints (10) where $\underline{B}^*_u$ denotes the optimal objective function value for $u = 0, \ldots, t\text{-}1$, ensure that optimization at the stage t does not spoil the objective function value achieved at the previous stages.

$$\sum_{j \in J} b_j h_{uj} \leq \underline{B}^*_u \quad for\ u = 0, \ldots t-1 \tag{10}$$

After the adjustments, the optimal solution of the lexicographic min-max problem can be obtained by step-by-step solving series of the following problems, where the t-th problem is formulated as follows.

$$Minimize \qquad \sum_{j \in J} b_j h_{tj} \tag{11}$$

$$Subject\ to\ (2) - (5)\ and\ (8) - (10)$$

An advantage of radial formulation consists in making use of the special structure of the dividing points where $0 = D_0 < D_1 < \ldots < D_v < D_{v+1}$. Let us determine the sequence $\{G_u\}$ so that we set $w = v$ and $G_u = D_{v+1-u}$. Under this adjustment, the right-hand-side of the constraints (8) can be reformulated as shown in (12).

$$\sum_{s=0}^{w} e_s x_{js} - G_{u+1} = \sum_{s=0}^{v} e_s x_{js} - D_{v-u} = \sum_{s=0}^{v} e_s x_{js} - \sum_{s=0}^{v-u+1} e_s \tag{12}$$

As the variable h_{uj} was originally defined by nonlinear expression (13), what includes both series of constraints (8) and (9), we obtain the substitution constraints (14) for h_{uj}.

$$h_{uj} = \max\left\{0, \sum_{s=0}^{w} e_s x_{js} - G_{u+1}\right\} \tag{13}$$

$$h_{uj} = \sum_{s=v-u}^{v} e_s x_{js} \tag{14}$$

Using the substitution (14), the constraints (10) can be adjusted to the constraints (15).

$$\sum_{j\in J} b_j \sum_{s=v-u}^{v} e_s x_{js} \le \underline{B}_u^* \quad for\ u = 0, \ldots t-1 \tag{15}$$

Now the optimal solution of the lexicographic min-max problem can be obtained by step-by-step solving series of the following problems for $t = 0, \ldots, v$.

$$Minimize \quad \sum_{j\in J} b_j \sum_{s=v-t}^{v} e_s x_{js} \tag{16}$$

Subject to (2) - (5) *and* (15).

4 Price of fairness

As mentioned in Section 1, the price of fairness was developed to evaluate the loss of the system criterion value caused by adopting the fair approach to the more affected users. The objective function value F computed as a sum of disutility values perceived by users can be expressed by (17) for the arbitrary solution $\mathbf{y}=[y_1, y_2, \ldots, y_m]$, where y_i takes the value of one, if a service solution is located at the possible location i and it takes the value of zero otherwise.

$$F(\mathbf{y}) = \sum_{j\in J} b_j \min\{d_{ij} : i \in I,\ y_i = 1\} \tag{17}$$

Let $\mathbf{y}^s$ and $\mathbf{y}^f$ denote system and fair optimal solutions respectively, then the relation between the system and fair optimum called the price of fairness [1] can be calculated by (18).

$$POF = \frac{F(\mathbf{y}^f) - F(\mathbf{y}^s)}{F(\mathbf{y}^s)} \tag{18}$$

To express the difference between two solutions of the PSSD concerning the different service center deployment, so called Hamming distance between the resulting vectors of location variables, e.g. $\mathbf{y}^s$ and $\mathbf{y}^f$ obtained for system and fair optimal design, can be used. The Hamming distance $H(\mathbf{y}^s, \mathbf{y}^f)$ represents "Manhattan distance" between two tops of unit hypercube and can be calculated by (19).

$$H(\mathbf{y}^f, \mathbf{y}^s) = \sum_{i=1}^{m} \left| y_i^f - y_i^s \right| \tag{19}$$

5 Computational study

The computational study is concentrated on the usage of radial formulation with homogenous system of radii defined by a series of the dividing points for both system and fair optimal public service system design using a commercial IP-solver. The approaches were tested on the pool of benchmarks obtained from the road network of Slovak Republic. The instances are organized so that they correspond to the administrative organization of Slovakia. For each self-governing region (Bratislava - BA, Banská Bystrica - BB, Košice - KE, Nitra - NR, Prešov - PO, Trenčín - TN, Trnava - TT and Žilina - ZA) all cities and villages with corresponding number of inhabitants b_j were taken. The coefficients b_j are rounded to hundreds. The number of possible service center locations $|I|$ is the same as the number of user locations $|J|$ in all solved instances and the network distance from user to the nearest located center was taken as an individual user´s disutility.

The experiments were organized so that the dividing points mentioned in Section 2 were determined accordingly to the procedure described in [10] at first. Using the radial approach to the PSSD with system criterion, the system optimal design of the public service system for given self-governing region was developed. For the obtained design given by vector $\mathbf{y}^s$ of location variables y_i the objective function *ObjF* was evaluated according to (17). The result evaluation contained also determination of the maximal disutility (distance) between a user and the nearest located service center. This value is denoted by *DmM* in all following tables**.** As we were able to compute the system optimal design using exact location-allocation formulation, we could compare the obtained approximate radial solution to the exact one. The *Gap* mentioned in the following tables represents the difference between the objective function these values expressed in percentage of the exact solution. The computational time of the radial approach in seconds is reported in the following tables in the columns denoted by *Time*. The computational time of the location-allocation approach is not presented here, but based on our previous experi-

ments published in [10], we note, that the radial model of the public service system design problem with system quality criterion performs approximately ten times faster.

Further, each instance was solved accordingly to the process for the radial approach to the PSSD with fair criterion. The same dividing points as ion the previous procedure were used and the problem (16), (2)-(5), (15) was solved and the fair optimal design of the public service system for given self-governing region was obtained. This design, which was also described by vector $\mathbf{y}^{\mathbf{f}}$ of location variables y_i, was evaluated in the following way. The objective function *ObjF*, *Time* and *DmM* were determined as in the system approach. In addition, we performed the comparison of the difference between the two vectors $\mathbf{y}^{\mathbf{s}}$ and $\mathbf{y}^{\mathbf{f}}$ so that we computed their Hamming distance according to (19) and plotted it in the tables in the column denoted by *HD*. In the column *BD*, there is reported number of users whose distance from the nearest service center equals to the maximal value *DmM* in the fair design. For the system design, we computed number of users whose distance from the nearest service center is greater than or equal to the same value *DmM* obtained from the fair design. This value is denoted by BF and the entries are placed in the part of table for system design. As mentioned above, the fair solution is paid by a loss of the average user's disutility. To express the loss, the price of fairness (*POF*) was evaluated accordingly to (18) and it was plotted in the column denoted by *POF*. All experiments were performed using the optimization software FICO Xpress 7.5 (64-bit, release 2013) for both location-allocation model and the covering approach with different quality criteria. The associated code was run on a PC equipped with the Intel® Core™ i7 2630QM processor with the parameters: 2.0 GHz and 8 GB RAM.

The experiments were performed with the data for each self-governing region of the Slovak Republic. The original problem comes from real emergency medical service system [4], [9], where given number of ambulances should be deployed in the given region so that ensure the rescue service for associated population. To obtain bigger pool of benchmarks for our computational study, the value of p was set in such a way, that the ratio of $|I|$ to p equals 2, 3, 4, 5, 10, 15, 20, 30, 40, 50 and 60 respectively. The associated results for the region of Žilina for all the generated instances, which differ in the parameter p, are plotted in the Table 1.

$\|I\|$	p	RADIAL - SYSTEM CRIT.					RADIAL - FAIR CRIT.					
		Time	*ObjF*	*Gap*	*DmM*	*BF*	*Time*	*ObjF*	*DmM*	*BD*	*HD*	*POF*
315	158	0.73	2444	0.00	13	227	2.08	6115	4	176	92	150.21
315	105	0.78	5594	0.00	17	216	2.45	19339	6	1	102	245.71
315	79	0.87	8430	0.00	20	224	3.15	19979	7	160	78	137.00
315	63	0.89	11125	0.00	20	288	5.35	24083	8	387	68	116.48
315	32	0.97	21075	0.38	20	149	6.10	29576	14	13	42	40.34
315	21	0.90	28548	0.00	26	185	7.67	41488	16	79	26	45.33
315	16	1.17	34405	0.00	33	161	13.99	55001	20	38	24	59.86
315	11	1.25	48143	2.23	33	36	64.88	76219	28	1	20	58.32
315	8	1.48	58842	0.30	42	27	30.14	83088	31	42	12	41.21
315	7	1.44	63865	0.00	42	2	29.00	76804	42	2	12	20.26
315	6	1.65	69572	0.00	42	10	34.98	85933	36	3	6	23.52

Table 1 Results of the experiments for the self-governing region of Žilina

Since the detailed results for other self-governing regions had similar characteristic as obtained for the region of Žilina, we report only selected instances for the other regions.

	$\|I\|$	p	RADIAL - SYSTEM CRIT.				RADIAL - FAIR CRIT.					
			Time	*ObjF*	*DmM*	*BF*	*Time*	*ObjF*	*DmM*	*BD*	*HD*	*POF*
BA	87	22	0.19	6698	15	187	0.40	18996	8	7	24	183.61
BB	515	35	1.93	24318	26	180	187.68	38967	16	27	54	60.24
KE	460	31	2.08	27681	25	232	185.19	44130	15	2	48	59.42
NR	350	35	1.17	22651	17	142	163.15	37370	13	3	56	64.98
PO	664	34	2.89	32985	25	75	55.55	66893	18	5	54	102.80
TN	276	19	1.14	21437	28	104	7.33	34167	16	10	26	59.38
TT	249	17	1.17	25425	24	167	134.43	36929	15	39	24	45.25
ZA	315	32	0.97	21075	20	149	6.10	29576	14	13	42	40.34

Table 2 Results of the experiments for the self-governing regions of Slovakia for selected parameter p

The value of parameter *p* in the selected instances was chosen so that it corresponds to the original set of problems from real emergency medical service system.

6 Conclusions

The main goal of this paper was to present and compare different approaches to the public service system design with different quality criteria. We presented an approximate approach based on radial formulation of the problem with homogenous system of radii given by dividing points. This approach was used for both system optimal public service system design and the fair one. We presented the solutions of several real instances and compared the resulting designs. Comparing the reported results in Table 1 and Table 2, we can conclude that the studied approach based on radial formulation with the dividing points can be successfully applied for the public service system design regardless which of the quality criteria is applied.

It can be noticed that the system and fair designs differ a lot, not only in the system criterion, what is the sum of network distances, but also in the characteristics of fairness. Considering the most impacted users from both designs, we can found that these users in the fair case have mostly their service center much nearer than the users from the system case. This observation holds especially for bigger values of parameter *p*. This dependency also corresponds to the Hamming distance between the pair of location vectors, where the Hamming distance decreases with decreasing value of the parameter *p*. A similar trend appears in the loss of system objective function value, when the fair criterion is accented. The better characteristics of the fairness obtained for relatively small values of *p* are paid by bigger price of fairness. At the end we can add that the accuracy of the approximate approaches is very good and the computational times are acceptable. Thus we can conclude that we presented useful tool for the middle-sized public service system design, which can be implemented using common commercial optimization software.

Acknowledgement

This work was supported by the research grants VEGA 1/0296/12 “Public Service Systems with Fair Access to Service”, VEGA 1/0339/13 “Advanced microscopic modelling and complex data sources for designing spatially large public service systems“, APVV-0760-11 “Designing of Fair Service Systems on Transportation Networks” and by the following project: University Science Park of the University of Žilina (ITMS: 26220220184) supported by the Research & Development Operational Program funded by the European Regional Development Fund.

References

[1] Bertsimas, D., Farias, V. F., Trichakis, N.: *The Price of Fairness*. Oper. Res. 59 (2011), pp. 17–31

[2] Current, J., Daskin, M., Schilling, D.: *Discrete network location models*. In: Drezner Z (ed) et al. Facility location. Applications and theory. Berlin: Springer, pp. 81-118, (2002).

[3] García, S., Labbé, M., Marín, A.: *Solving large p-median problems with a radius formulation*. In: INFORMS Journal on Computing 23 (4), pp. 546-556, (2011).

[4] Ingolfsson, A., Budge, S., Erkut, E.: *Optimal ambulance location with random delays and travel times*, In: Health Care management science, 2008, 11 (3), pp. 262-274, (2008).

[5] Janáček, J.: *Approximate Covering Models of Location Problems*. In: Lecture Notes in Management Science: Proceedings of the 1st International Conference ICAOR ´08, Vol. 1, Sept. 2008, Yerevan, Armenia, pp. 53-61, ISSN 2008-0050, (2008).

[6] Janáček, J., Kvet, M.: *Approximate solving of large p-median problems*. In: ORP3: proceedings, Cádiz, Spain, September 13-17, 2011, Cádiz: 2011, ISBN 978-84-9828-348-8, pp. 221-225, (2011).

[7] Janáček, J., Kvet, M.: *Public service system design with disutility relevance estimation*. In: Proceedings of the 31 st international conference ”Mathematical Methods in Economics”, September 11-13, 2013, Jihlava, Czech Republic, pp 332 – 337

[8] Janáček, J., Linda, B., Ritschelová, I.: *Optimization of Municipalities with Extended Competence Selection*. In: Prager Economic Papers-Quarterly Journal of Economic Theory and Policy, Vol. 19, No 1, 2010, pp. 21-34, (2010).

[9] Jánošíková, L.: *Emergency Medical Service Planning*. In: Communications –Scientific Letters of the University of Žilina, Vol. 9, No 2, pp. 64-68, (2007).

[10] Kvet, M., Janáček, J.: *Impact of population distribution on the optimal public service system design*. In Quantitative Methods in Economics 2014, Bratislava: Ekonóm, ISBN 978-80-225-3868-8, pp. 146-154

[11] Marianov, V., Serra, D.: *Location problems in the public sector*. In: Drezner Z (ed.) et al. Facility location. Applications and theory. Berlin: Springer, pp. 119-150, (2002).

[12] Ogryczak, W., Sliwinski, T.: *On Direct Methods for Lexicographic Min-Max Optimization*. In M. Gavrilova et al. (Eds.): ICCSA 2006, LNCS 3982, pp. 802-811, Springer-Verlag Berlin Heidelberg, 2006.

On functional definition of time-series models

Petr Lachout[1]

Abstract. We present a discussion on a class of econometric models for time series given implicitly as a solution of a system of functional equations. Particularly, processes AR, ARMA, ARCH, GARCH are models of such a structure. These nonlinear time series models are treated from several points of view. Task of existence of a solution, possibility of numerical simulations and description of solutions are considered and partially solved in the article.

Keywords: nonlinear time series, functional equation, ARMA process, ARCH process, recurrent procedure.

JEL classification: C22
AMS classification: 91B84; 62M10

1 Introduction

Linear models for time series are based on properties of the Hilbert space L_2 and the calculus of polynomials in the backward shift operator, see e.g. [1], [2], [4], [9]. We will consider more complex models for time series written by a system of functional equations; see [3], [5], [6], [7], [8]. This theoretical description and consequent properties possess relevant impact to practice. Modern analysis of financial and economic data can be hardly made without this fruitful setup.

Our realization is focused to time series models based on a system of functional equations where the functions are Lipschitz continuous. Controlling Lipschitz constant we derive partial response to some natural tasks. We discuss existence of a solution, possibility of numerical simulation and description of solutions.

The observations are illustrated on two particular models. The first model is covering AR model and the second is closed to GARCH model.

2 Description of the model

We assume that we observe a realization of a random real vector series $\mathsf{Y} = (\mathsf{Y}(t), t \in \mathbb{Z})$, where for each time $t \in \mathbb{Z}$ we have $\mathsf{Y}(t) = (\mathsf{Y}_1(t), \mathsf{Y}_2(t), \dots, \mathsf{Y}_J(t))^\top$. We suppose the random process Y is fulfilling an equality

$$\forall\, t \in \mathbb{Z} \qquad \mathsf{Y}(t) = \mathsf{f}(\mathsf{Y}(t-1), \mathsf{Y}(t-2), \dots, \mathsf{Y}(t-p); \mathsf{Z}(t), \mathsf{Z}(t-1), \dots, \mathsf{Z}(t-q)), \tag{1}$$

where $J \in \mathbb{N}$, $K \in \mathbb{N}$, $p \in \mathbb{N}$, $q \in \mathbb{N}_0$, $\mathsf{f} : \mathbb{R}^{p\times J} \times \mathbb{R}^{(q+1)\times K} \to \mathbb{R}^J$ is a measurable function, $\mathsf{Z} = (\mathsf{Z}(t), t \in \mathbb{Z})$, where for each time $t \in \mathbb{Z}$ we have $\mathsf{Z}(t) = (\mathsf{Z}_1(t), \mathsf{Z}_2(t), \dots, \mathsf{Z}_K(t))^\top$.

We are interested in several questions: if the equation (1) possesses a solution, how to make simulations of a solution, to find a description of the solution, and, if there is a solution having some particular properties.

To abbreviate forthcoming formulae, we introduce a notation

$$\begin{aligned} \mathsf{Y}(t \circledcirc k) &= \mathsf{Y}(t), \mathsf{Y}(t-1), \dots, \mathsf{Y}(t-k), \\ \mathsf{Z}(t \circledcirc k) &= \mathsf{Z}(t), \mathsf{Z}(t-1), \dots, \mathsf{Z}(t-k). \end{aligned} \tag{2}$$

Hence, the relation (1) converts to an abbreviate form

$$\mathsf{Y}(t) = \mathsf{f}(\mathsf{Y}(t-1 \circledcirc p-1); \mathsf{Z}(t \circledcirc q)). \tag{3}$$

[1]Dept. Probability and Statistics, Charles University in Praha, Faculty of Mathematics and Physics, Petr.Lachout@mff.cuni.cz

3 Main results

3.1 Forward solution

Applying recurrent plug-in, we define a sequence of functions:

$$\mathsf{f}_1\left(\mathsf{y}\left(t-1\circledcirc p-1\right);\mathsf{e}\left(t\circledcirc q\right)\right)=\mathsf{f}\left(\mathsf{y}\left(t-1\circledcirc p-1\right);\mathsf{e}\left(t\circledcirc q\right)\right), \tag{4}$$

$$\forall\ k\in\mathbb{N}\qquad \mathsf{f}_{k+1}\left(\mathsf{y}\left(t-k-1\circledcirc p-1\right);\mathsf{e}\left(t\circledcirc q+k\right)\right)= \tag{5}$$

$$=\mathsf{f}_k\left(\mathsf{f}\left(\mathsf{y}\left(t-k-1\circledcirc p-1\right);\mathsf{e}\left(t-k\circledcirc q\right)\right),\mathsf{y}\left(t-k-1\circledcirc p-2\right);\mathsf{e}\left(t\circledcirc q+k-1\right)\right).$$

Hence, we are receiving

$$\forall\ t\in\mathbb{Z}\quad\forall\ k\in\mathbb{N}\quad \mathsf{Y}\left(t\right)\ =\ \mathsf{f}_k\left(\mathsf{Y}\left(t-k\circledcirc p-1\right);\mathsf{Z}\left(t\circledcirc q+k-1\right)\right), \tag{6}$$

Consequently, for each starting time $\tau\in\mathbb{Z}$ and each starting sequence $\mathsf{y}\left(\tau\right),\mathsf{y}\left(\tau-1\right),\ldots,\mathsf{y}\left(\tau-p+1\right)$, we are receiving a random process defined for all $t\in\mathbb{Z}$, $t>\tau$

$$\widehat{\mathsf{Y}}\left(t\,|\,\tau;\mathsf{y}\left(\tau\circledcirc p-1\right)\right)\ =\ \mathsf{f}_t\left(\mathsf{y}\left(\tau\circledcirc p-1\right);\mathsf{Z}\left(t\circledcirc q+t-1\right)\right). \tag{7}$$

Immediately, we have the first observation.

Proposition 1. *For each starting time $\tau\in\mathbb{Z}$ and starting sequence $\mathsf{y}\left(\tau\right),\mathsf{y}\left(\tau-1\right),\ldots,\mathsf{y}\left(\tau-p+1\right)$, the random process $\left(\widehat{\mathsf{Y}}\left(t\,|\,\tau;\mathsf{y}\left(\tau\circledcirc p-1\right)\right),t>\tau\right)$ is always fulfilling the property (1) for all $t\in\mathbb{Z}$, $t>\tau$.*

The form (7) is a recipe how to produce a numerical solution of (1) valid from a given starting time.

3.2 Lipschitz constants

In some cases dependence on the starting sequence is vanishing with time growth. We will show such a case controlling Lipschitz constant of f. Let us state a definition of Lipschitz constant for a function $F:\mathcal{X}\times\mathcal{Z}\to\mathcal{Y}$ where $\mathcal{X},\mathcal{Y},\mathcal{Z}$ are Banach spaces:

$$\mathsf{Lip}\left(F\right)\ =\ \left\{\frac{\|F(\eta;z)-F(\xi;z)\|}{\|\eta-\xi\|}\ :\ \eta,\xi\in\mathcal{X},\ \eta\neq\xi,\ z\in\mathcal{Z}\right\}. \tag{8}$$

Now, we derive an estimation of Lipschitz constants for our system of functions f_k, $k\in\mathbb{N}$.

Let us consider starting time $\tau=0$ and two starting sequences $\eta\left(0\right),\eta\left(-1\right),\ldots,\eta\left(-p+1\right)$ and $\xi\left(0\right),\xi\left(-1\right),\ldots,\xi\left(-p+1\right)$. Using recurrently (3), we determine further members of these sequences, i.e. for each $t\in\mathbb{N}$ we set

$$\begin{aligned}\eta\left(t\right)\ &=\ \mathsf{f}\left(\eta\left(t-1\circledcirc p-1\right);\mathsf{Z}\left(t\circledcirc p\right)\right),\\ \xi\left(t\right)\ &=\ \mathsf{f}\left(\xi\left(t-1\circledcirc p-1\right);\mathsf{Z}\left(t\circledcirc p\right)\right).\end{aligned}$$

Now, we estimate Euclid distance between members of these sequences. Actually, we will estimate square of Euclid distance, because it is easier for writing. Let us consider $t\in\mathbb{N}$ and estimate:

$$\begin{aligned}\left\|\eta\left(t\right)-\xi\left(t\right)\right\|^2\ &=\ \left\|\mathsf{f}\left(\eta\left(t-1\circledcirc p-1\right);\mathsf{Z}\left(t\circledcirc p\right)\right)-\mathsf{f}\left(\xi\left(t-1\circledcirc p-1\right);\mathsf{Z}\left(t\circledcirc p\right)\right)\right\|^2 \qquad (9)\\ &\leq\ \mathsf{Lip}\left(\mathsf{f}\right)^2\left\|\eta\left(t-1\circledcirc p-1\right)-\xi\left(t-1\circledcirc p-1\right)\right\|^2\\ &\leq\ \mathsf{Lip}\left(\mathsf{f}\right)^2\sum_{k=1}^{p}\left\|\eta\left(t-k\right)-\xi\left(t-k\right)\right\|^2.\end{aligned}$$

Now, we proceed with estimation for the sum. The formula (9) is recurrently plugged in for members

with the highest argument.

$$\begin{aligned}
&\sum_{k=1}^{p} \|\eta(t-k)-\xi(t-k)\|^2 \leq \\
&\quad\leq \sum_{k=2}^{p}\left(1+\mathsf{Lip}(\mathsf{f})^2\right)\|\eta(t-k)-\xi(t-k)\|^2 + \\
&\quad\quad +\sum_{k=1}^{1}\mathsf{Lip}(\mathsf{f})^2\|\eta(t-p-k)-\xi(t-p-k)\|^2 \\
&\quad\leq \sum_{k=2}^{p}\left(1+\mathsf{Lip}(\mathsf{f})^2\right)\|\eta(t-k)-\xi(t-k)\|^2 + \\
&\quad\quad +\sum_{k=1}^{1}\left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)-1\right]\|\eta(t-p-k)-\xi(t-p-k)\|^2 \\
&\quad\leq \sum_{k=3}^{p}\left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)+\mathsf{Lip}(\mathsf{f})^2\left(1+\mathsf{Lip}(\mathsf{f})^2\right)\right]\|\eta(t-k)-\xi(t-k)\|^2 + \\
&\quad\quad +\sum_{k=1}^{2}\left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)+\mathsf{Lip}(\mathsf{f})^2\left(1+\mathsf{Lip}(\mathsf{f})^2\right)-\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^{k-1}\right]\cdot \\
&\quad\quad\quad \cdot\|\eta(t-p-k)-\xi(t-p-k)\|^2 \\
&\quad\leq \sum_{k=3}^{p}\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^2\|\eta(t-k)-\xi(t-k)\|^2 + \\
&\quad\quad +\sum_{k=1}^{2}\left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^2-\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^{k-1}\right]\|\eta(t-p-k)-\xi(t-p-k)\|^2 \\
&\quad\leq \cdots \leq \\
&\quad\leq \sum_{k=l+1}^{p}\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^l\|\eta(t-k)-\xi(t-k)\|^2 + \qquad (10)\\
&\quad\quad +\sum_{k=1}^{l}\left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^l-\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^{k-1}\right]\|\eta(t-p-k)-\xi(t-p-k)\|^2 \\
&\quad\leq \cdots \leq \\
&\quad\leq \sum_{k=1}^{p}\left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^p-\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^{k-1}\right]\|\eta(t-p-k)-\xi(t-p-k)\|^2 \\
&\quad\leq \left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^p-1\right]\sum_{k=1}^{p}\|\eta(t-p-k)-\xi(t-p-k)\|^2. \qquad (11)
\end{aligned}$$

Using (9) and recurrently (10), (11), we are receiving following estimate of Lipschitz constants for our system of functions f_k, $k\in\mathbb{N}$.

$$\begin{aligned}
\mathsf{Lip}(\mathsf{f}_1) &= \mathsf{Lip}(\mathsf{f}), \qquad (12)\\
\mathsf{Lip}(\mathsf{f}_{np+k+1}) &\leq \mathsf{Lip}(\mathsf{f})\left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^p-1\right]^{n/2}\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^{k/2} \\
&\quad \forall\, n\in\mathbb{N}_0,\ k=0,1,\ldots,p-1.
\end{aligned}$$

In the case

$$\mathsf{Lip}(\mathsf{f})<\sqrt{2^{1/p}-1}, \qquad (13)$$

we simplify form of estimation to

$$\mathsf{Lip}(\mathsf{f}_k) \leq \mathsf{Lip}(\mathsf{f})\left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^p-1\right]^{\frac{k-p}{2p}}\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^{(p-1)/2} \quad \forall\, k\in\mathbb{N}. \qquad (14)$$

3.3 Solution properties

Applying the estimate (14), we can estimate distance between two solutions of (1).

Theorem 2. *Let* Y *be fulfilling (1),* $\tau \in \mathbb{Z}$ *and* $\mathsf{y}(\tau), \mathsf{y}(\tau-1), \ldots, \mathsf{y}(\tau-p+1)$ *be a starting sequence. If* $\mathsf{Lip}(\mathsf{f}) < \sqrt{2^{1/p}-1}$ *then always for each* $t \in \mathbb{Z}$, $t > \tau$

$$\begin{aligned} &\left\| \mathsf{Y}(t) - \widehat{\mathsf{Y}}(t \,|\, \tau; \mathsf{y}(\tau \circledcirc p-1)) \right\| \leq \\ \leq\ & \mathsf{Lip}(\mathsf{f}) \left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^p - 1 \right]^{\frac{t-\tau-p}{2p}} \left(1+\mathsf{Lip}(\mathsf{f})^2\right)^{(p-1)/2} \left\| \mathsf{Y}(0 \circledcirc p-1) - \mathsf{y}(0 \circledcirc p-1) \right\|. \end{aligned} \tag{15}$$

Particularly, we have

$$\lim_{t \to +\infty} \left\| \mathsf{Y}(t) - \widehat{\mathsf{Y}}(t \,|\, \tau; \mathsf{y}(\tau \circledcirc p-1)) \right\| = 0. \tag{16}$$

This means that distance among solutions of (1) is vanishing exponentially if time is increasing. Also, distance between a solution of (1) and numerical solution (7) is vanishing exponentially, moreover, independently on the starting sequence.

3.4 Particular solution

Theorem 3. *Let* Y *be fulfilling (1),* $\tau \in \mathbb{Z}$ *and* $\mathsf{Lip}(\mathsf{f}) < \sqrt{2^{1/p}-1}$. *Then for each* $t \in \mathbb{Z}$, $t > \tau$

$$\begin{aligned} &\left\| \mathsf{Y}(t) - \widehat{\mathsf{Y}}(t \,|\, \tau; 0, 0, \ldots, 0) \right\| \leq \\ \leq\ & \mathsf{Lip}(\mathsf{f}) \left[\left(1+\mathsf{Lip}(\mathsf{f})^2\right)^p - 1 \right]^{\frac{t-\tau-p}{2p}} \left(1+\mathsf{Lip}(\mathsf{f})^2\right)^{(p-1)/2} \left\| \mathsf{Y}(\tau \circledcirc p-1) \right\|. \end{aligned} \tag{17}$$

If, moreover, there is a constant $\mathsf{C} \in \mathbb{R}$ *such that* $\mathsf{E}\left[\|\mathsf{Y}(s)\|^2\right] \leq \mathsf{C}$ *for all* $s \in \mathbb{Z}$, $s < 0$ *then*

$$\left\| \mathsf{Y}(t) - \widehat{\mathsf{Y}}(t \,|\, \tau; 0, 0, \ldots, 0) \right\| \to 0 \quad \textit{in probability and in } \mathsf{L}_2 \textit{ if } \tau \to -\infty. \tag{18}$$

Thus,

$$\widehat{\mathsf{Y}}(t \,|\, \tau; 0, 0, \ldots, 0) \to \mathsf{Y}(t) \quad \textit{in probability and in } \mathsf{L}_2 \textit{ if } \tau \to -\infty. \tag{19}$$

The observation is interesting. It is saying to us that a solution of (1) with uniformly bounded second moments can be constructed in L_2 as a limit (19).

Note, if $\mathsf{Z} = (\mathsf{Z}(t), t \in \mathbb{Z})$ are i.i.d. then the construction (19) is revealing that random process Y is (strictly) stationary.

4 Examples

In this section we present two examples to illustrate our observations.

4.1 Linear model

Firstly, we consider a linear function f. We choose the simplest case

$$\forall\, t \in \mathbb{Z} \qquad \mathsf{Y}(t) = \phi \mathsf{Y}(t-1) + \mathsf{Z}(t). \tag{20}$$

Then, the parameters are

$$J = 1,\ K = 1,\ p = 1,\ q = 0,\ \mathsf{f}(y\,; z) = \phi y + z,\ \mathsf{Lip}(\mathsf{f}) = |\phi|. \tag{21}$$

For each starting time $\tau \in \mathbb{Z}$ and starting value $y \in \mathbb{R}$. we receive for each $t \in \mathbb{Z}$, $t > \tau$

$$\widehat{\mathsf{Y}}\left(t \,|\, \tau; y\right) \quad = \quad \sum_{k=0}^{t-\tau} \phi^k \mathsf{Z}\left(t-k\right) + \phi^{t-\tau+1} y. \tag{22}$$

Theorems 2, 3 are saying for (20) the following.

Let Y be fulfilling (20), $\tau \in \mathbb{Z}$ and $|\phi| < 1$. Then

$$\lim_{t \to +\infty} \left\| \mathsf{Y}\left(t\right) - \left(\sum_{k=0}^{t-\tau} \phi^k \mathsf{Z}\left(t-k\right) + \phi^{t-\tau+1} y \right) \right\| = 0 \tag{23}$$

with a rate $|\phi|^t$.

If, moreover, there is a constant $\mathsf{C} \in \mathbb{R}$ such that $\mathsf{E}\left[\left\| \mathsf{Y}\left(s\right) \right\|^2 \right] \leq \mathsf{C}$ for all $s \in \mathbb{Z}$, $s < 0$ then

$$\sum_{k=0}^{+\infty} \phi^k \mathsf{Z}\left(t-k\right) \quad \text{is sumable in probability and in } \mathsf{L}_2, \tag{24}$$

$$\mathsf{Y}\left(t\right) = \sum_{k=0}^{+\infty} \phi^k \mathsf{Z}\left(t-k\right) \text{ a.s.} \tag{25}$$

If $\mathsf{Z} = \left(\mathsf{Z}\left(t\right), t \in \mathbb{Z}\right)$ are i.i.d. then this simple case converts to AR(1) process.

4.2 Varying volatility model

Here we consider model with varying volatility.

$$\begin{aligned} \forall\, t \in \mathbb{Z} \qquad & \mathsf{Y}_1\left(t\right) = \mathsf{Y}_2\left(t-1\right) \mathsf{Z}_1\left(t\right), \\ & \mathsf{Y}_2\left(t\right) = \sqrt{\alpha_0 + \alpha_1 \mathsf{Y}_2\left(t-1\right)^2 + \mathsf{Z}_2\left(t\right)^2}, \end{aligned} \tag{26}$$

where $\alpha_0 > 0$, $\alpha_1 > 0$, $\Delta > 0$ and $|\mathsf{Z}_1\left(t\right)| \leq \Delta$.

Then, the parameters are

$$J = 2,\, K = 2,\, p = 1,\, q = 0,\, \mathsf{f}\left(y\,;z\right) = \left(\frac{y_2 z_1}{\sqrt{\alpha_0 + \alpha_1 y_2 + z_2^2}} \right),\, \mathsf{Lip}\left(\mathsf{f}\right) \leq \sqrt{\alpha_1 + \Delta^2}. \tag{27}$$

For each starting time $\tau \in \mathbb{Z}$ and starting value $y \in \mathbb{R}^2$. we receive for each $t \in \mathbb{Z}$, $t > \tau$

$$\widehat{\mathsf{Y}}_1\left(t \,|\, \tau; y_1, y_2\right) \quad = \quad \sqrt{ \frac{\alpha_0}{1-\alpha_1} + \sum_{k=0}^{t-\tau-2} \alpha_1^k \mathsf{Z}_2\left(t-k-1\right)^2 + \alpha_1^{t-\tau-1} \left(y_2 - \frac{\alpha_0}{1-\alpha_1} \right) } \, \mathsf{Z}_1\left(t\right), \tag{28}$$

$$\widehat{\mathsf{Y}}_2\left(t \,|\, \tau; y_1, y_2\right) \quad = \quad \sqrt{ \frac{\alpha_0}{1-\alpha_1} + \sum_{k=0}^{t-\tau-1} \alpha_1^k \mathsf{Z}_2\left(t-k\right)^2 + \alpha_1^{t-\tau} \left(y_2 - \frac{\alpha_0}{1-\alpha_1} \right) }. \tag{29}$$

Theorems 2, 3 are saying for (26) the following.

Let Y be fulfilling (26), $\tau \in \mathbb{Z}$ and $\alpha_1 + \Delta^2 < 1$. Then

$$\lim_{t \to +\infty} \left\| \mathsf{Y}_1\left(t\right) - \sqrt{ \frac{\alpha_0}{1-\alpha_1} + \sum_{k=0}^{t-\tau-2} \alpha_1^k \mathsf{Z}_2\left(t-k-1\right)^2 + \alpha_1^{t-\tau-1} \left(y_2 - \frac{\alpha_0}{1-\alpha_1} \right) } \, \mathsf{Z}_1\left(t\right) \right\| = 0, \tag{30}$$

$$\lim_{t \to +\infty} \left\| \mathsf{Y}_2\left(t\right) - \sqrt{ \frac{\alpha_0}{1-\alpha_1} + \sum_{k=0}^{t-\tau-1} \alpha_1^k \mathsf{Z}_2\left(t-k\right)^2 + \alpha_1^{t-\tau} \left(y_2 - \frac{\alpha_0}{1-\alpha_1} \right) } \right\| = 0. \tag{31}$$

with an exponential rate $(\alpha_1 + \Delta^2)^{t/2}$.

If, moreover, there is a constant $\mathsf{C} \in \mathbb{R}$ such that $\mathsf{E}\left[\left\|\mathsf{Y}(s)\right\|^2\right] \leq \mathsf{C}$ for all $s \in \mathbb{Z}$, $s < 0$ then

$$\sum_{k=0}^{+\infty} \alpha_1^k \mathsf{Z}_2(t-k)^2 \quad \text{is sumable in probability and in } \mathsf{L}_1, \tag{32}$$

$$\mathsf{Y}_1(t) = \sqrt{\frac{\alpha_0}{1-\alpha_1} + \sum_{k=0}^{+\infty} \alpha_1^k \mathsf{Z}_2(t-k-1)^2}\, \mathsf{Z}_1(t) \text{ a.s.} \tag{33}$$

$$\mathsf{Y}_2(t) = \sqrt{\frac{\alpha_0}{1-\alpha_1} + \sum_{k=0}^{+\infty} \alpha_1^k \mathsf{Z}_2(t-k)^2} \text{ a.s.} \tag{34}$$

Consequently, if $\mathsf{Z} = (\mathsf{Z}(t), t \in \mathbb{Z})$ are i.i.d. then the construction (33), (34) is revealing that random process Y is (strictly) stationary.

5 Conclusion

Models of the form (1) fulfilling (13) possesses several solutions valid from a given starting time. If a solution has uniformly bounded second moments then we can construct it in L_2 using (19). If the noise is i.i.d. then the solution is (strictly) stationary.

Acknowledgements

The research was supported by the Czech Science Foundation (grant P402/12/G097).

References

[1] Box, G.E.P.; Jenkins, G.M.: *Time Series Analysis Forecasting and Control.* Holden-Day, San Francisco, 1976.

[2] Brockwell, P.J.; Davis, R.A.: *Time Series: Theory and Methods* Springer-Verlag, New York, 1987.

[3] Dhrymes, Phoebus J.: *Topics in Advanced Econometrics II - Linear and Nonlinear Simultaneous Equations.* Springer-Verlag, Berlin, 1994.

[4] Eds. Frank, J.; Härdle, W.; Martin, D.: *Robust and Nonlinear Time-Series Analysis.* Springer-Verlag, New York, 1983.

[5] Hommes, C.: Financial markets as nonlinear adaptive evolutionary systems. Quantitative Finance **1,1 (2001)**, 149-167.

[6] Masry, Elias; Tjøstheim, Dag: Nonparametric estimation and identification of nonlinear ARCH time series. Econometric Theory **11,1 (1995)**, 258-289.

[7] Priestley, M.B.: *Non-Linear and Non-Stationary Time Series.* Academic Press, London, 1988.

[8] Tjøstheim, D.: Non-Linear Time Series and Markov Chains. Advances in Applied Probability **22,3 (1990)**, 587-611.

[9] Wei, W.W.S.: *Time Series Analysis.* Addison-Wesley Publishing Company, Inc., Redwood City, 1994.

New tools for complementing the *h*-index: an empirical study

Tommaso Lando[1], Lucio Bertoli-Barsotti[2]

Abstract. The *h*-index, introduced by Hirsch in 2005, is used by major citation databases to evaluate the academic performance of a researcher. The advantages and disadvantages of the *h*-index have been studied and discussed so far and many authors proposed other well known variants. In this paper we present a new general class of bibliometric indices based on the quality and the quantity of the most relevant publications. We compare some of the obtained indices with the decisions of a Committee of five professors concerning a group of Physicists which applied for the Italian National Scientific Qualification (the so called "Abilitazione Scientifica Nazionale, the new process to recruit University Professors in Italy, based on scientific qualification criteria). Many well known indices, such as the *A*-index or the *e*-index, are sensitive to few highly cited publications. Conversely, our approach, based on a sum of increasing and concave functions, is aimed to reward researchers with a more stable and reliable scientific production. The empirical results show that the new class of indices effectively increases the capacity of discrimination among researcher with a similar value of the *h*-index.

Keywords: *h*-index, *h*-type indices, concave functions.

JEL Classification: C43, D70, D80
AMS Classification: 62P25, 62G10

1 Introduction

Bibliometric indices are used to evaluate researchers basing on their scientific productivity and impact on the scientific community. Among these indices the most widely used is surely the *h*-index [6] which measures the academic performance of an author simply basing on his/her number of publications and corresponding citations. Actually, the *h*-index is defined as the maximum number of papers *h* that received a number of citations greater or equal to *h*. The advantages and disadvantages of the *h*-index have been studied and discussed so far. Many authors proposed variants of the *h*-index introducing new variables which should be taken into account, such as the number of co-authors ([6]; [12]), the self-citations [10], the age of the papers [7] or the age of the author [1]. Other authors focused on other drawbacks of the *h*-index and proposed alternatives based just on publications and citations, without introducing additional variables. Within this framework, in this paper we express our point of view and propose an alternative class of bibliometric indices which is aimed to improve the *h*-index in terms of its accuracy and sensitivity to the "form" of the citations distribution.

The main success of the *h*-index is probably due to its simplicity and its robustness, in that it is insensitive to low-impact publications with few or no citations. On the other hand, the *h*-index is also insensitive to highly cited publications: as soon as such a publication is part of the *h*-core (the group of the *h* most highly cited papers; [9]), its actual number of citations has no longer influence. Moreover, the only number *h* seems to be too poor and "rough" to discriminate among authors with a similar scientific production. This problem is known as the "low resolution" of the Hirsch index: actually it is quite common to find researchers with equal *h*-index (in particular for normal/ low values of the *h*-index). For all these reasons, many authors proposed alternatives to the *h*-index based on publications and citations: the *g*-index [4] is sensitive to exceptional publications; the *A*-index, the *R*-index [7] and the *e*-index [12] complement the *h*-index by measuring the overall citation "intensity" in the *h*-core. Conversely, it seems that all these alternatives can be influenced too easily by few outstanding highly cited papers. Thus they actually measure the *impact* of an author on the scientific community rather than scientific productivity. Impact (or magnitude) measures seem to be appropriate for the evaluation of high level scientists (Nobel prizes, medalists…) but could be misleading in other contexts. Suppose to refer to the common case when the evaluation of a researcher is aimed at the qualification for an academic position, i.e., assistant/associate professor etc. Among applicants of the same level, we believe that a bibliometric index should reward the more regular researchers, in order to address research institutions towards reliable selections. In a bibliometric context, a sort of risk-averse attitude suggests to choose, between two researchers of the same level, the one with a more uniform or regular distribution of citations. Hence, to measure the overall level and the reliability of a researcher, in section 2 we propose to use a sum of non-decreasing and concave functions, evaluated on the set of the most relevant papers. Note that concave functions reduce the effect of extremely cited publications, which is not a

[1] VŠB TU, Department of Finance 70121, Sokolská 33 Ostrava (Czech Republic), tommaso.lando@vsb.cz
[2] University of Bergamo, via dei Caniana 2, 24127 Bergamo (Italy), lucio.bertoli-barsotti@unibg.it

common tendency in the literature. Then, in section 3, the proposed indices are tested and compared on a dataset consisting of 149 Physicists which applied for the Italian National Scientific Qualification (ASN).

2 Methodology

For a given researcher (say "researcher a") with a total number of publications $n(a)$ let us denote with $c_i(a)$ the number of citations of paper *i* ($i = 1, \dots, n(a)$), and let the papers be ranked in decreasing order of the number of citations that they received, so that $c_1(a) \geq c_2(a) \geq \cdots \geq c_{n(a)}(a)$. We refer to the vector $c(a) = (c_1(a), c_2(a), \dots, c_{n(a)}(a))$ as the *citation distribution* of researcher *a*. The *h*-index of researcher *a*, say $h(a)$, can be defined as follows:

$$h(a) = max\{i \mid c_i(a) \geq i\}.$$

The number *h* identifies a set of relevant papers, the so called *h-core*. It is interesting to observe that the Hirsch index mainly depends on the form of the citation distribution, rather than the number of citations: for a given total number of citations $C(a)$, *h* is greater when the distribution is "squared" and smaller when the distribution has a "L" form. In particular *h* is maximized by $\min\{c_1(a), n(a), \lfloor\sqrt{C(a)}\rfloor\}$ where $[x]$ denotes the integer part of the real number *x* and n is the number of papers with at least 1 citation ([2]; [3]). In the common situation when $\lfloor\sqrt{C(a)}\rfloor \leq c_1(a), n(a)$ the maximum value of *h* is actually $\lfloor\sqrt{C(a)}\rfloor$, which means that $c_i(a) \geq \lfloor\sqrt{C(a)}\rfloor$ for $i = 1, \dots, \lfloor\sqrt{C(a)}\rfloor$ thus the distribution of citations is basically a square with side $\lfloor\sqrt{C(a)}\rfloor$. One of the main alternatives to the *h*-index is the *g*-index, proposed by Egghe [4]: g is the largest number such that the most cited g papers received, together, at least g^2 citations. Then *g*-index of researcher *a* can be defined as:

$$g(a) = max\{i \mid C_i(a) \geq i^2\},$$

where $C_i(a) = \sum_{j=1}^{i} c_j(a)$. Similarly to *h*, the number g identifies a set of relevant papers, the so called *g-core* (note that this set exists only if the researcher has a number of publications $n \geq g$). It is interesting to note that $c_i(a) \geq i$ for $i = 1, \dots, h$ yields $C_i(a)/i \geq i$ for $i = 1, \dots, h$, thus, by definition, the *h*-core is a subset of the *g*-core (and $g \geq h$). The *g*-index is sensitive to highly cited publications and it does not strictly depend on the form of the distribution. Actually, unlike *h*, g can be maximized both by a "squared" or a "L" distribution, as proved in the sequel. Define by $I(x)$ the function such that $I(x) = 1$ if $x \geq 0$ and $I(x) = 0$ otherwise. Assume to fix the number of total citation to given value, say C. Denote by *l* an author with an "L" distribution defined as $c(l) = (C, 0, \dots, 0)$. For any author *a* with C citations observe that:

$$\max_a g(a) = g(l) = \sum_{i=1}^{n} I\left(\lfloor\sqrt{C}\rfloor - i\right) = \lfloor\sqrt{C}\rfloor = h(s) > h(l) = 1,$$

where s is a "squared form" distribution, such that $c_i(s) \geq \lfloor\sqrt{C}\rfloor$ for $i = 1, \dots, \lfloor\sqrt{C}\rfloor$. Nevertheless, consider that, for s, we obtain $C_j(s) = \sum_{j=1}^{i} c_i(s) \geq i\lfloor\sqrt{C}\rfloor$, thus

$$g(s) \geq \sum_{i=1}^{n} I(i\lfloor\sqrt{C}\rfloor - i^2) = \sum_{i=1}^{n} I\left(\lfloor\sqrt{C}\rfloor - i\right) = \lfloor\sqrt{C}\rfloor.$$

We conclude that $\max_a g(a) = g(s) = g(l) = h(s) = \lfloor\sqrt{C}\rfloor$. Definitely, it seems that both indices agree when the citation distribution is squared, which happens when a researcher produce a relevant number of good quality publications, rather than a few outstanding ones. In this paper we base on this concept and propose a class of bibliometric indices aimed to measure the overall quality of the most relevant publications based on the form of the distribution. Speaking of which, the following indices have been proposed to complement the *h*-index by measuring the quality and intensity of publications within the *h*-core: the *A*-index is defined as the average citations in the *h*-core: $A(a) = \frac{1}{h(a)}\sum_{i=1}^{h(a)} c_i(a)$; the *R*-index is simply the square root of the total citations within the core $R(a) = \sqrt{\sum_{i=1}^{h(a)} c_i(a)}$; the *e*-index is the square root of the sum of the excess citations in the *h*-core, so $e^2(a) = \sum_{i=1}^{h(a)}(c_i(a) - h(a)) = \sum_{i=1}^{h(a)} c_i(a) - (h(a))^2$; henceforth we will consider, for simplicity, $E(a) = e^2(a)$, say the *E*-index. The authors who propose those indices suggest to discriminate among scientists with equal *h*-index by considering the pairs (*h*, *A*), (*h*, *R*) or (*h*, *E*). Nevertheless, the considered indices (*A*, *R* and *E*) are strictly related between them and are essentially based on a sum, which, from a statistical and mathematical point of view, poorly describes the distribution of the citations within the core.

If we take the *E*-index as a benchmark we can imagine two researchers with equal *h* value but also equal sum of excess citations. Take, for instance, researcher a and researcher b, with $h(a) = h(b) = 10$ and with a distribution of excess citations (within the *h*-core) respectively given by:

Researcher a: (10,10,10,10,10,10,10,10,10,10)
Researcher b: (100,0,0,0,0,0,0,0,0,0)

Now, the question is: basing on this information, who is the best researcher? (Note that $E(a) = \sum_{i=1}^{10}(c_i(a) - 10) = E(b) = \sum_{i=1}^{10}(c_i(b) - 10) = 100$, $A(a) = A(b)$ and $R(a) = R(b)$). Our opinion is that researcher *a* should be preferred to researcher *b*, as the overall quality of the publications in the core should not depend too strictly on only one or few publications. In general we believe that a researcher with a more "even" or "homogeneous" distribution of excess citations should be preferred. The same result should hold, a fortiori, if the overall citations of researcher *a* is higher than the overall citations of researcher *b* (i.e. $E(a) > E(b)$). This could be said by writing

$$\sum_{i=0}^{k} \delta_{h-i}(a) \geq \sum_{i=0}^{k} \delta_{h-i}(b), \text{ for every } k=0,1,2,\ldots,h\text{-}1$$

where $h = h(a) = h(b)$ and

$$\delta(a) = \big(\delta_1(a), \ldots, \delta_h(a)\big) = (c_1(a) - h, \ldots, c_h(a) - h);$$
$$\delta(b) = \big(\delta_1(b), \ldots, \delta_h(b)\big) = (c_1(b) - h, \ldots, c_h(b) - h).$$

It is interesting to note that this corresponds to the notion of *"weak" majorization* [8]. More specifically, $\delta(a)$ is said to be weakly majorized from above by $\delta(b)$, and we write $\delta(a) \prec^w \delta(b)$.
Note that in the very special case in which

$$\sum_{i=1}^{h} \delta_i(a) = \sum_{i=1}^{h} \delta_i(b) \text{ (i.e. } (E(a) = E(b)),$$

(as in our first example) $\delta(a)$ is said to be (strongly) *majorized* by $\delta(b)$, and we write $\delta(a) \prec \delta(b)$. Majorization [5] is a pre-order defined on vectors with the same number of components. It is aimed to determine whether the components of a vector are more (or less) "spread out" or more (or less) "equal" than the components of another vector. In our situation $\delta(a) \prec^w \delta(b)$ means that, among researchers *a* and *b*, researcher *a* is more "stable" from a qualitative point of view, as its papers (starting from the less cited within the core) have a cumulative sum of citations which is always higher compared to researcher *b*. For this reason researcher *a* should be more reliable than researcher *b*. Thus, we search for indices that preserve weak (and strong) majorization. The desired index should follow the following formula

$$E_\phi(a) = \sum_{i=1}^{h(a)} \phi\big(\delta_i(a)\big) = \sum_{i=1}^{h(a)} \phi(c_i(a) - h(a)),$$

where ϕ is an increasing and concave function. Indeed, the following property is proved in [8] (p. 109):

Property 1. *$\delta(a) \prec^w \delta(b)$ if and only if $E_\phi(a) \geq E_\phi(b)$ for any increasing and concave function ϕ.*

Hence, by taking for example $\phi(t) = t^\alpha$ $(0 \leq \alpha \leq 1)$, we define the following family of bibliometric indicators:

$$E_\alpha(X) = \sum_{i=1}^{h(a)} (\delta_i(a))^\alpha = \sum_{i=1}^{h(a)} (c_i(a) - h(a))^\alpha.$$

Notice that $E_1(a) = E(a)$ and $E_0(a) = h(a)$ (with the assumption $0^0 = 1$) so the defined class generalizes both the *E*- and *h*-indices. In particular $h(a) \leq E_\alpha(a) \leq E(a)$. The choice of α may vary depending on how strongly we desire to reduce the weight of highly cited papers. Speaking of which, we believe that highly cited papers should not be overrated: those papers can increase their citations quite easily, due to the popularity of the journal, the popularity from the previous citations etc. (Matthew effect) and there is no need to emphasize them further. In the next section the proposed class of indices are tested on real data.

3 Case of study - Results

The Italian National Scientific Qualification (Abilitazione Scientifica Nazionale, ASN) is a new process to recruit University Professors in Italy, based on scientific qualification criteria. In this paper we focused on the set of 149 Physicists that were applicants in the 2012 ASN for a full professorship position in Condensed Matter

Physic. An expert panel of evaluators (a Committee of 5 members) was asked, by the Italian University Ministry, to approve ("habilitate") or to reject each candidate. Let us call such a dichotomous variable "habilitation", with values 0 (= rejected applicant) and 1 (= approved applicant). In Italy, habilitation is a requirement to be eligible for full professorship.

The complete list of publications and corresponding citations for each of these applicants were retrieved by us from Scopus in January 2014. From the original sample of 149 datasets (for almost all the candidates for full professorship the status was that of "Associate Professor"; the list of candidates was retrieved from the URL: http://abilitazione.miur.it/public/index.php), 18 datasets were discarded from the analyses due to insufficient citation data (e.g. an *h*-index less than 2) or difficulties in identifying the scientist. Then, for each one of the selected 131 datasets, several different research productivity indices were computed, such as *h*-index, *g*-index, *R*-index, *e*-index, *A*-index, as well as other basic metrics such as total number of papers, total number of cited papers, total number of citations, maximum number of citation for a single paper, etc.. We also computed the E_α-index, for α equal to 0.25, 0.50, 0.75 and 1.

Table 1 shows the Pearson correlation coefficient for each pair of indices considered in our study. Interestingly enough, for every fixed alpha, the E_α-index was always more correlated with the *g*-index (the largest Pearson correlation coefficient was found between *g* and E_α for α=0.5) than with the *h*-index. Then the index $E_{0.5}$ seems to be a good proxy for the *g* index.

	h	***g***	***R***	***e***	***A***	$E_{0.25}$	$E_{0.50}$	$E_{0.75}$
g	0.910							
R	0.895	0.999						
e	0.809	0.979	0.986					
A	0.682	0.920	0.934	0.976				
$E_{0.25}$	0.971	0.965	0.956	0.903	0.798			
$E_{0.50}$	0.903	0.983	0.980	0.960	0.887	0.976		
$E_{0.75}$	0.805	0.963	0.966	0.977	0.943	0.909	0.976	
$E_{1.00}$	0.691	0.908	0.916	0.950	0.954	0.812	0.910	0.977

Table 1 Observed Pearson correlation coefficients

We considered the problem of comparing the values of these indicators for rejected and approved applicants. The main idea is that the value of a (good) research productivity index of approved applicant can be expected to have an higher value than that given by the same indicator for a rejected applicant. As expected, in the whole sample (131 cases, of which 69% approved by the Committee), the *h*-index was related to the judgement of the Committee: indeed, the average *h*-index of rejected applicants was 14.3, while the average *h*-index of approved applicants was 24.9. Moreover, i) of the 21 scientists with an *h*-index less than 13 only 5 were approved; ii) of the 21 scientists with an *h*-index greater than 29, no one was rejected. But what happens for fixed *h*? to respond to this question, the 89 scientists with an *h*-index $12<h<30$ were grouped into 5 classes. As a general rule, as can be seen in Table 2, the percentage of approved applicants, as expected, grows with the value of *h*. Moreover, it results that for *every fixed* "level" of the *h*-index, the E_α-index (α=0.25) of the approved applicants are, on average, higher than the E_α-index of the rejected applicants. This is also illustrated in Figure 1.

h	E_α **– rej**	E_α **– app**	**% of approved**
13-16	27.86 (6)	34.53 (2)	25%
17-19	34.94 (7)	38.05 (6)	46%
20-21	41.50 (5)	42.38 (9)	64%
22-25	47.16 (4)	49.26 (28)	88%
26-29	56.65 (2)	61.36 (20)	91%

Table 2 Averaged values of E_α-index for approved and rejected applicants for different values of their *h* index (α=0.25). In parentheses the number of cases. (E_α – rej = average E_α-index for the rejected candidates; E_α – app = average E_α-index for the approved candidates)

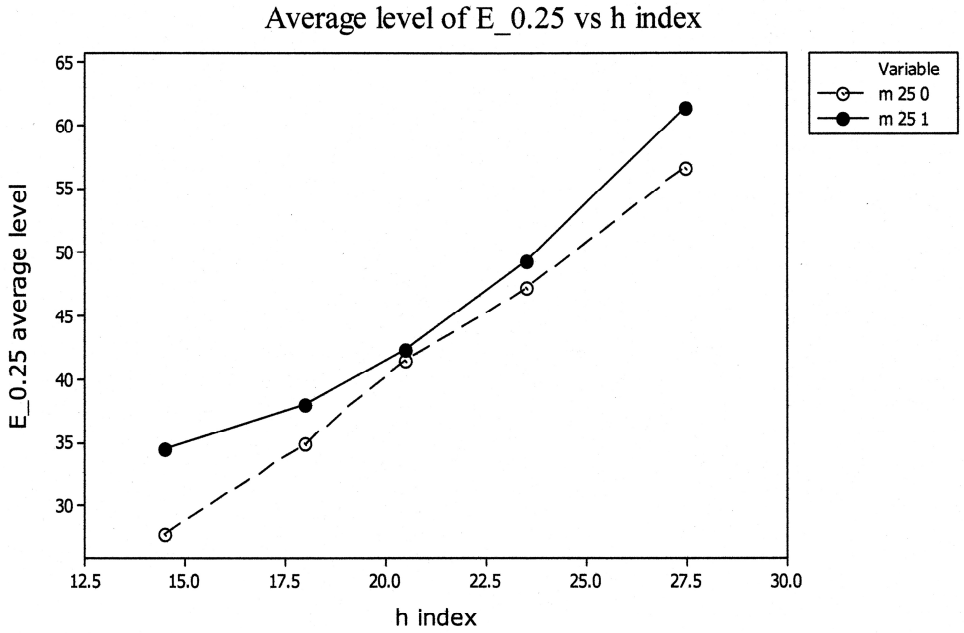


Figure 1 Average values of E_α-index for approved and rejected applicants for different values of their *h* index (α=0.25).

Due to the high percentage of approved and rejected applicants in the areas with respectively high and low value of *h*, all the considered indices result appropriate to "explain" the habilitation variable. In order to find some significant differences among those indices it was necessary to focus on the core of the dataset, selecting those researchers which are more "uncertain" with regard to the outcome of the evaluation. Hence, to better understand if specific relationship exists between the E_α-index and the rejected/approved status for a given level of the *h*-index, we isolated a more specific subsample of scientists with similar values of the *h*-index. This sample included all and only those with an *h*-index between 15 and 21. This set of authors was composed by 32 applicants; 15 of these were rejected, and 17 were approved (thus the percentages of approved and rejected are really close, reflecting the uncertainty of the evaluation outcome within this group).

index	rejected applicants group	approved applicants group	*p-value*
h	15, 16, 16, 17, 17, 18, 18, 18, 19, 19, 20, 20, 20, 21, 21	15, 16, 18, 18, 19, 19, 19, 19, 20, 20, 20, 20, 20, 20, 20, 21, 21	*0.113*
g	24, 25, 26, 27, 28, 29, 30, 32, 34, 34, 35, 36, 37, 40, 40	26, 26, 28, 29, 31, 31, 32, 33, 35, 36, 36, 40, 41, 43, 44, 49, 50	*0.070*
$E_{0.25}$	27.39, 27.57, 31.05, 32.13, 34.41, 34.44, 34.82, 36.56, 37.45, 37.70, 38.18, 39.24, 39.53, 44.47, 47.89	29.07, 29.83, 32.06, 35.09, 35.42, 37.01, 37.45, 40.01, 40.14, 40.44, 42.86, 43.39, 43.73, 47.56, 47.74, 48.48, 48.48	*0.041*
$E_{0.50}$	50.46, 55.17, 63.92, 67.59, 68.13, 73.50, 77.63, 78.28, 86.24, 87.34, 94.25, 96.33, 96.66, 100.75, 122.76	51.39, 59.14, 67.16, 69.35, 75.69, 82.30, 86.58, 89.92, 96.83, 101.87, 108.60, 114.84, 121.04, 121.22, 126.64, 127.19, 143.20	*0.048*
$E_{0.75}$	99.06, 117.84, 138.42, 145.10, 148.44, 163.85, 167.70, 190.65, 214.80, 221.71, 239.48, 254.79, 261.94, 295.98, 333.39	91.58, 132.96, 145.56, 146.24, 175.07, 188.80, 195.14, 214.52, 250.53, 253.86, 287.68, 328.87, 342.62, 369.96, 381.72, 477.58, 489.39	*0.087*
$E_{1.00}$	205, 264, 314, 327, 339, 375, 379, 496, 580, 590, 610, 762, 771, 949, 1050	168, 321, 325, 329, 434, 449, 460, 540, 661, 687, 806, 974, 1120, 1192, 1294, 1881, 2147	*0.093*

Table 3 Equal distribution test (alternative: less than). *P-values* for different indices.

Assuming the independence of these two samples, to test the null hypothesis of an equal distribution (vs alternative hypothesis that the sample of the approved applicants dominates, at the first order, the sample of the rejected applicants) a Wilcoxon-Mann-Whitney nonparametric test was done on the ranked data (see Table 3). As can be seen in Table 3, the E_α-index (e.g. for α=0.25) allows to reject the null hypothesis, while the *h*-index does not (note that the *e*-index preserves the ranking of $E_{1.00}$-index, then it gives the same p-value 0.093, not significant).

4 Conclusions

The E_α-index measures in some way the citation intensity in the *h* core, but following the principle that a more "homogeneous" distribution of excess citations in the *h* core should be preferred to a skewed distribution. Like other research productivity indices, such as the *R*-index, or Zhang's *e*-index (we recall that the square of *e*-index is equal to the E_α-index for α=1), E_α is not suited to be used as a stand-alone index, but as a complement to the *h*-index. This study gives strong empirical evidence that the E_α-index is useful for discriminating between authors with similar *h*-indices and then for ranking researchers conditionally on the value of the *h*-index.

Acknowledgements

This paper has been elaborated in the framework of the project Opportunity for young researchers, reg. no. CZ.1.07/2.3.00/30.0016, supported by Operational Programme Education for Competitiveness and co-financed by the European Social Fund and the state budget of the Czech Republic.

References

[1] Abt, H., A.: A publication index that is independent of age. *Scientometrics* 91 (3) (2012), 863-868.
[2] Anderson, T., R.,
[3] Bartolucci, F.: On a possible decomposition of the *h*-index. *Journal of the American Society for Information Science and Technology* 63 (10) (2012), 2126-2127.
[4] Bertol-Barsotti, L.: Improving a decomposition of the *h*-index. *Journal of the American Society for Information Science and Technology* 64 (7) (2013), 1522.
[5] Egghe, L.: An improvement of the h-index: The g-index. *ISSI Newsletter* 2 (1) (2006), 8-9.
[6] Hardy G. H., Littlewood J. E., Pòlya G.: *Inequalities.* Cambridge University Press (1934).
[7] Hirsch, J., RE: An index to quantify an individual's scientific research output. *Proceedings of the national Academy of Sciences of the United States of America* 102 (46) (2005), 16569-16572.
[8] Jin, B., Liang, L., Rousseau, R., Egghe, L.: The R- and AR-indices: complementing the h-index. *Chinese Science Bulletin*, 52 (6) (2007), 885-863.
[9] Marshall, A., Olkin, I.: *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York, (1979).
[10] Rousseau, R.: New developments related to the Hirsch index. *Science Focus* 1 (2006), 23-25
[11] Schreiber, M.: Self-citations correction for the Hirsch Index. *Europhysics Letters* 78 (3) (2007), Article Number: 30002.
[12] Schreiber, M.: A case study of the Modified Hirsch Index hm accounting for multiple coauthors. *Journal of the American Society for Information Science and Technology* 60 (6) (2009), 1272-1282.
[13] Zhang, C.: The e-index, complementing the h-index for excess citations. *PLoS ONE* 4 (5) (2009), Article Number: e5429.

Life Expectancy and Modal Age at Death in the Czech Republic in 1920-2012

Jana Langhamrová[1], Markéta Arltová[2]

Abstract. The aim of this paper is to show changes and trends in the evolution of life expectancy and modal age of death in the Czech Republic in years 1920-2012 using the model life tables. A model of life tables will be calculated and constructed on the basis of data of deaths and population in middle of year published by the CZSO. Trends in secondary life expectancy and modal age at death of men and women at selected ages with special attention to male excess mortality will be assessed. Mortality at the highest ages will be settled by using the Gompertz-Makeham function. Modal age a death will be calculated based on the parameters of this function, more precisely. The evaluation will be also likely for age of an *x*-year-old person. Life expectancy is characteristic of the type of average, it is influenced by infant mortality and death rates in middle age, modal age at death is characteristic of type of modus and it is influenced by trends in mortality at older ages. On the basis of trends in the life expectancy and modal age of death and indicator characteristics in terms of statistics it will be proposed, which of this characteristics is now more appropriate in the evaluation process of demographic aging and longevity rating.

Keywords: life expectancy, modal age at death, normal length of life, male excess mortality, life tables, time series, Czech Republic.

JEL Classification: J11, C22
AMS Classification: 26E60, 37M10

1 Introduction

At present, the majority of developed countries deal with the phenomenon of population ageing. This process is accompanied by an increasing rate of elderly persons, especially in economically developed countries. Mortality rates are improving and people live longer. The development of mortality in developed countries was neither linear nor logistic. Periods of faster or slower decline varied in time and some trends in mortality were unexpected. Currently, decrease in mortality in older age groups is mentioned as a major factor in the aging population.

In studies of longevity the most often used indicators are life expectancy and modal age at death. Length of life is usually expressed by the indicator of life expectancy at age *x*. The values of life expectancy and modal age at death are different from the view of time evolution. This is particularly because life expectancy is the average age of deceased persons in the stationary population, whereas modal age at death is the most common age at death.

In this paper will not be described in detail characteristics in terms of demographic trends and influences that have caused changes in their development. We focus here more on methods of calculation of these characteristics and their interaction in terms of time series. Below will be analysed relationship between individual time series from the viewpoint if they are related to common stochastic trend or not.

This paper focuses on life expectancy at birth, life expectancy for 65-year-old person and modal age at death for men and women in the Czech Republic in 1920-2012.

2 Calculation of life expectancy and modal age at death

Life expectancy (e_x) is the average number of years to be lived by an x-year-old person while maintaining the mortality like in the pursued period. It is an indicator which shows mortality rates for all age groups. For calcu-

[1] University of Economics, Department of Statistics and Probability, W. Churchill Sq. 4, 130 67 Prague, Czech Republic, xlanj18@vse.cz.
[2] University of Economics, Department of Statistics and Probability, W. Churchill Sq. 4, 130 67 Prague, Czech Republic, arltova@vse.cz

lating of life expectancy is needed to create so called life tables. Their detailed description will not be placed here, only key calculations will be shown below.

The life expectancy of a person at the exact age x can be described as the ratio of the number of remaining years of life (T_x) and the number of survivors to exact age (l_x):

$$e_x = \frac{T_x}{l_x}. \tag{1}$$

The most used ages of calculation of life expectancy are 0 years and 65 years. It is a sort of length of whole life for 0-year-old person. And for 65-year-old person it means the number of years that has person in front of him.

The so called modified Gompertz-Makeham function is usual to calculate life expectancy for highest ages, which will be explained below.

Modal age at death (e_N) can be estimated as a rough estimate (age at last birthday with the maximum number of deaths). We are looking for an age when the number of deaths in mortality tables is the highest.

However, as already mentioned, this is only a rough estimate and more accurate results are obtained when using parameters of Gompertz-Makeham function. Calculation of modal age at death was performed by Fiala [6]. For this calculation it is necessary to know some source data as total deaths is specified age group ($M_{t,x}$) and mid-year population in the same age group ($\overline{S_{t,x}}$) or total population at the beginning of one year ($S_{t,x}$) or total population at the end of one year ($S_{t+1,x}$). Primary characteristics of mortality are age-specific death rates. Age-specific death rate for one calendar year is calculated using the formula

$$m_{t,x} = \frac{M_{t,x}}{\overline{S}_{t,x}}. \tag{2}$$

When we don´t know the number of mid-year population, but we have the number of total population at the beginning of the year t and $t+1$, we use the following formula

$$m_{t,x} = \frac{M_{t,x}}{\frac{S_{t,x} + S_{t+1,x}}{2}}, \tag{3}$$

where $M_{t,x}$ is the number of total deaths in completed years x and in calendar year t, $S_{t,x}$ is the total population aged x at the beginning of the year t. Compensation of age-specific death rates at age 60 and above can be calculated by Gompertz-Makeham equation

$$\tilde{m}_x^{(GM)} = a + b \cdot c^{x+\frac{1}{2}}. \tag{4}$$

We select the beginning of the first $x_0 = 60$ and the length of intervals $k = 8$, where k we understand as the interval. We calculate the summation of empirical death rates by age in each interval and mark them as G_1, G_2, G_3

$$G_1 = \sum_{x=60}^{67} m_x, \; G_2 = \sum_{x=68}^{75} m_x, \; G_3 = \sum_{x=76}^{83} m_x. \tag{5}$$

Now we can calculate the value of the parameter c of Gompertz-Makeham function whose eighth squared value can be expressed by using the sum of empirical death rates by age in each interval

$$c^8 = \frac{G_3 - G_2}{G_2 - G_1}. \tag{6}$$

Furthermore, it is necessary to calculate the value of the subexpression, by which we can express the remaining two parameters of the function

$$K_c = c^{60,5} \cdot (1 + c + ... + c^7) = c^{60,5} \cdot \frac{c^8 - 1}{c - 1}. \tag{7}$$

We can calculate the parameters b by using next expressions

$$b = \frac{G_2 - G_1}{K_c \cdot (c^8 - 1)}, \; a = \frac{G_1 - b \cdot K_c}{8} \tag{8}$$

And according to these parameters of the Gompertz-Makeham function we can now calculate the modal age at death more precisely

$$\hat{y} = \frac{\ln \frac{\ln c - 2a + \sqrt{(\ln c - 4a) \cdot \ln c}}{2b}}{\ln c}. \tag{9}$$

In 1999 Koschin [12] published his modified Gompertz-Makeham function. In this modified model, there is a new parameter γ which provides a decrease in speed of increases of mortality rate with age. Modified Gompertz-Makeham function can be expressed as

$$\tilde{m}_x^{(GM)} = a + b * c^{x_0 + \frac{1}{\gamma} * \ln[\gamma * (x - x_0) + 1]}. \tag{10}$$

2.1 Consequences between this characteristics from the viewpoint of time series

Analysis of the relationship between the integrated time series is meaningful only if these time series are cointegrated, i.e. they are related to common stochastic trend. If time series are not cointegrated, we can analyze them with regression and there is a status so called spurious regression [8]. We perform the differentiation of true regression from spurious regression by using Engle-Granger cointegration test [5]. Results of this test include Table 2.

		women		**men**	
		t_{ADF}	**Prob.**	t_{ADF}	**Prob.**
*e*0	level	3.741501	0.9999	-3.090536	0.1148
	1st diff.	-9.248715	0.0000	-11.54967	0.0000
*e*65	level	1.314515	0.9986	1.675978	0.9767
	1st diff.	-13.38346	0.0000	-13.73936	0.0000
*e*N	level	0.365407	0.9804	1.539657	0.9690
	1st diff.	-13.54865	0.0000	-12.13429	0.0000

Table 1 Unit Root Test of the Time Series and of the First Differences *Source: own calculations*

		women				**men**			
		tau-stat.	Prob.*	z-stat.	Prob.*	tau-stat.	Prob.*	z-stat.	Prob.*
*e*0 & *e*N	E0	-1.1976	0.8594	-3.5560	0.8394	-2.1724	0.4423	-3.3167	0.8562
	EN	-0.6267	0.9519	-2.0429	0.9296	-0.4521	0.9664	-1.3214	0.9580
*e*65 & *e*N	E65	-4.4664	0.0026	-30.305	0.0032	-3.8721	0.0151	-26.349	0.0090
	EN	-4.6683	0.0013	-31.273	0.0025	-4.0139	0.0101	-27.269	0.0071

Table 2 Engle-Granger test of cointegration *Source: own calculations*

It is obvious that the cointegration relationship (it is tested hypothesis that time series are cointegrated) can only be found in the time series E65 and EN both women and men. In the case of time series E0 and EN is a specious relationship.

Dependent variable: EN Sample (adjusted): 1921 2012				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19.25274	3.946546	4.878377	0.0000
EN(-1)	0.665326	0.069280	9.603388	0.0000
E65	1.278005	0.103171	12.38722	0.0000
E65(-1)	-0.739978	0.153579	-4.818236	0.0000
R-squared	0.991994		Durbin-Watson stat	2.270807
F-statistic	3634.523		Prob(F-statistic)	0.000000
Diagnostics tests			Statistics	Prob.
Breusch-Godfrey Serial Correlation LM Test:			1.935956	0.1505
Jarque-Bera Test			0.785017	0.6753
ARCH Test			2.212939	0.1404

Table 3 Model ADL - women *Source: own calculations*

Dependent variable: EN Sample (adjusted): 1921 2012				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.43179	3.819949	3.778006	0.0003
ENM(-1)	0.720812	0.072396	9.956517	0.0000
E65M	1.517349	0.161090	9.419277	0.0000
E65M(-1)	-0.955382	0.210648	-4.535442	0.0000
R-squared	0.964544		Durbin-Watson stat	1.989476
F-statistic	797.9733		Prob(F-statistic)	0.000000
Diagnostics tests			Statistics	Prob.
Breusch-Godfrey Serial Correlation LM Test:			0.463373	0.6307
Jarque-Bera Test			26.52597	0.0000
ARCH Test			0.644413	0.4243

Table 4 Model ADL - men *Source: own calculations*

Weak exogeneity test found that both men and women in the E65 model is weakly exogenous and EN is endogenous variable. So we used for the analysis of interrelationships the ADL model (Autoregressive Distributed Lag) [9], which is usually expressed in the form of ADL (1,1)

$$Y_t = c + \alpha_1 Y_{t-1} + \beta_1 X_t + \beta_2 X_{t-1} + a_t, \tag{11}$$

and then from the ADL model derived ECM (Error Correction Model) [5]

$$\Delta Y_t = c + \beta_1 \Delta X_t + \gamma(Y_{t-1} - \beta X_{t-1}) + a_t, \tag{12}$$

where $\beta = (\beta_1 + \beta_2)/(1 - \alpha_1)$ and $\gamma = \alpha_1 - 1$, which differentiate the short-term and long-term relationships between time series.

The resulting ADL model of dependence of EN on E65 for women can be written as

$$\widehat{eN}_t = 19.2527 + 0.6653\ eN_{t-1} + 1.278\ e65_t - 0.7399\ e65_{t-1}, \tag{13}$$

and for men

$$\widehat{eN}_t = 14.4318 + 0.7208\ eN_{t-1} + 1.5173\ e65_t - 0.9554\ e65_{t-1.} \tag{14}$$

And from these relations were derived EC models, first for women

$$\Delta\widehat{eN}_t = 19.2527 + 1.278\ \Delta e65_t - 0.3347(eN_{t-1} - 1{,}608 e65_{t-1}). \tag{15}$$

and next for men

$$\Delta\widehat{eN}_t = 14.4318 + 1.5173\ \Delta e65_t - 0.2792(eN_{t-1} - 2{,}013 e65_{t-1}). \tag{16}$$

We obtain through parameters $\gamma^W = 0.3347$ and $\gamma^M = 0.2792$ from these models and information about the speed, which systems respond to tilting the balance and the value of long-term multiplier $\beta^W = 1.608$ and $\beta^M = 2.013$.

3 Development of life expectancy and modal age of death

It is visible from the Figure 1, that during whole watched period have women higher life expectancy at birth than men. This is caused by male excess mortality during their live, in simple terms men have higher intensity of mortality than women. This difference is visible for next two characteristics too (Figure 2 and 3), life expectancy for 65-year-old person and modal age at death. These differences are given by biology, but not in such extend like in the Czech Republic in 1920-2012. Differences were not so high at the beginning of the period, but during 50th occurs to increase of this distinction due to worsening of mortality condition of men. From 90th are mortality issues for men getting better.

By comparing the evolution of life expectancy at birth, life expectancy at age 65 and modal age at death, it is visible that modal age at death is not increasing as rapidly as life expectancy at age *x*. A similar trend both men and women shows Figures 2 and 3, which we have confirmed by the results of the cointegration relationship between time series.

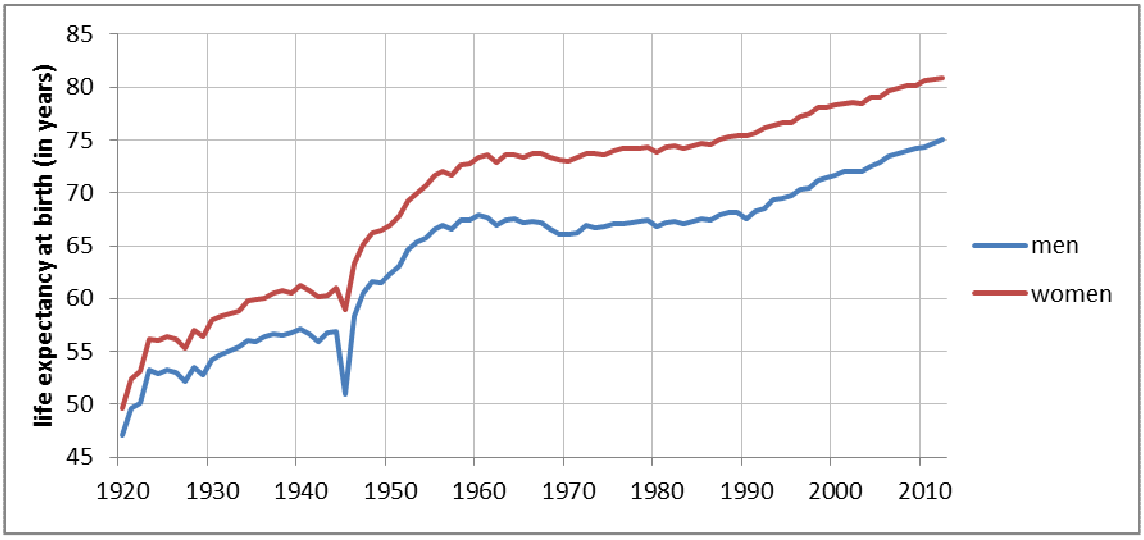


Figure 1 Life expectancy at birth of men and women in the Czech Republic in 1920-2012

Source: data CZSO, own calculations

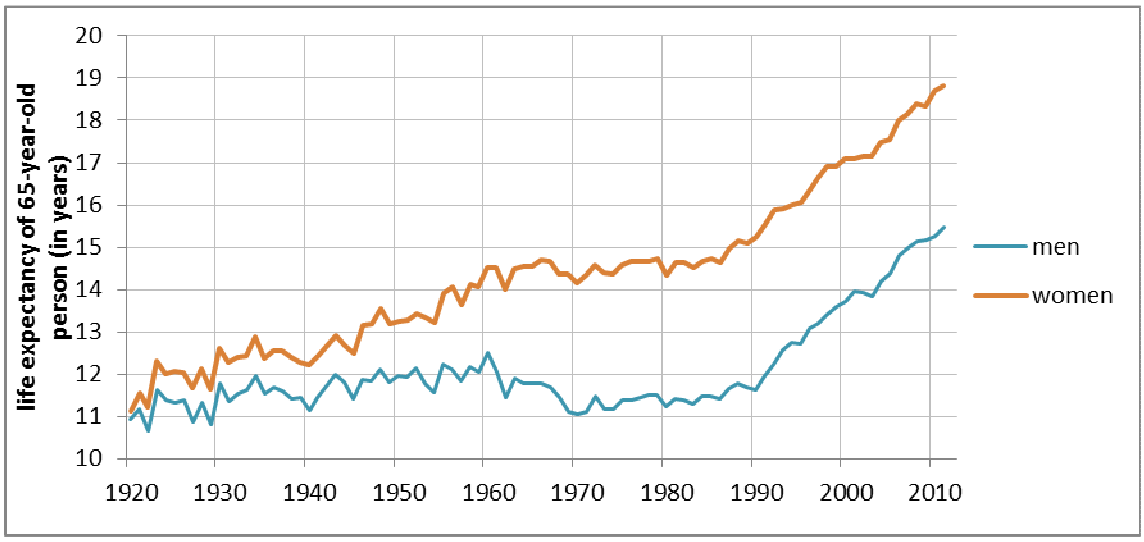


Figure 2 Life expectancy 65-years-old men and women in the Czech Republic in 1920-2012

Source: data CZSO, own calculations

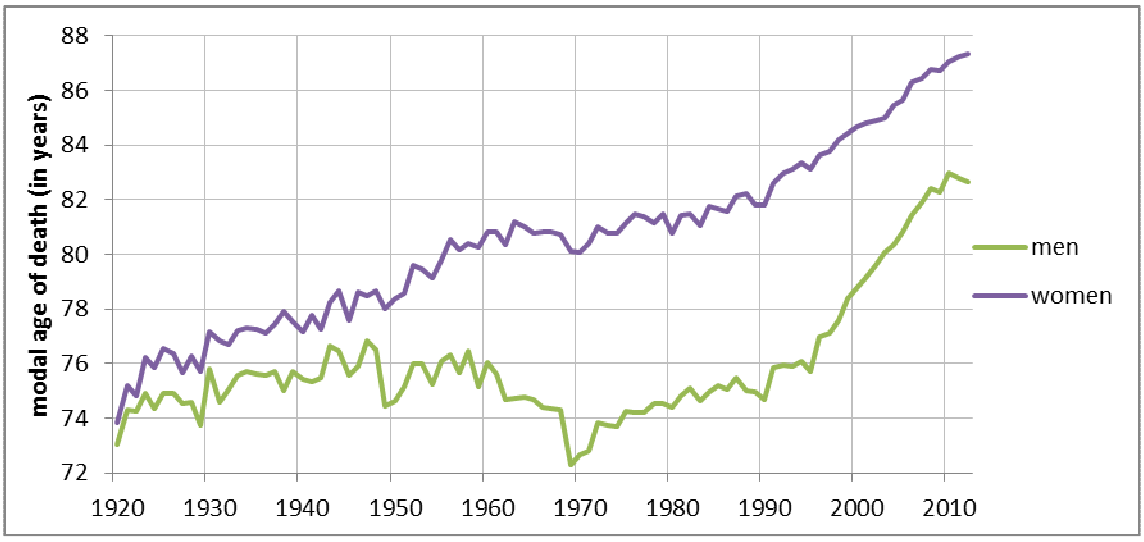


Figure 3 Modal age of death of men and women in the Czech Republic in 1920-2012

Source: data CZSO, own calculations

4 Conclusion

In this paper, we explored characteristics of life expectancy of *x*-year-old person and modal age of death. Life expectancy is characteristic of the type of average, it is influenced by infant mortality and death rates in middle age, modal age at death is characteristic of type of modus and it is influenced by trends in mortality at older ages. Time series of these characteristics were not compared from the demographical point of view, but in terms of the calculation and time series relations.

It is possible to prepare ADL (1,1) models of dependence of modal age at death on life expectancy for 65-year-old person for both sexes and derived EC models for both sexes from results of time series analysis. We obtain through parameters information about the speed, which systems respond to tilting the balance and the value of long-term from these models.

Characteristic of life expectancy is type of average and it is affected by extremes. There is although a relationship between life expectancy for 65-year-old person and modal age at death. It is not appropriate to use only this model instead of life expectancy because these relationships in time series are highly influenced by the chosen period, and many other factors. This model can describe the relations and interactions between the time series, but can not fully substitute the explanatory power of length of life indicator.

Acknowledgements

This paper was written with the support of Grant Agency of the Czech Republic No. P402/12/G097 DYME Dynamic Models in Economics and IGS Project No. 24/2013 Mortality and Aging of Population in the Czech Republic.

References

[1] Arlt, J., Arltová, M. (2009): *Ekonomické časové řady*. Professional Publishing.
[2] Breusch, T., Godfrey, L. G. (1986): Data Transformation Tests, *Economic Journal*, 96 (Souplement), 47 - 58.
[3] Darnell, A. C. (1994): *A Dictionary of Econometrics*, Edward Elgar.
[4] Dickey, D. A., Fuller, W. A. (1979): "*Distribution of Estimators for Autoregressive Time Series with a Unit Root."* Journal of American Statistical Association, Vol. 74, No. 366, pp. 427-431.
[5] Engle, R.F., Granger, C.W.J. (1987): *Cointegration and Error Correction Representation:* Estimation and Testing,. Econometrica, 55: 251-276.
[6] Fiala, T. *Výpočty aktuárské demografie v tabulkovém procesoru*. Praha: Oeconomica, 2005. 177 s. ISBN 80-245-0821-4.
[7] Gavrilov, L.A. and Heuveline, P. (2003). *Aging of Population.* In: Demeny, P. and McNicoll, G. (eds.). The Encyclopedia of Population. New York: Macmillan Reference USA: 32-37
[8] Granger, C. W. J., Newbold, P. (1974): *Spurious Regressions in Econometrics*. Journal of Econometrics, Vol. 2, No. 2, pp. 111-120.
[9] Hendry, D., Pagan, A., Sargan, J. (1984): *Dynamic Specifications*. Chapter 18 in Handbook of Econometrics, Vol II (ed., Z. Griliches and M. Intriligator), North Holland
[10] Jarque, C., Bera, A. (1980): *Efficient Tests for Normality, Heteroscedasticity, and Serial Independence of Regression Residuals*, Economics Letters, 6, 255 - 259
[11] Kannisto V. "*Mode and Dispersion of the Length of Life*" Population: An English Selection. 2001;13:159–71
[12] Koschin, F.: *Jak vysoká je intenzita úmrtnosti na konci lidského života?* Demografie, 1999, roč. 41, č. 2. s. 105-119.
[13] Wilmoth John R. "*Demography of Longevity: Past, Present and Future Trends*" Experimental Gerontology. 2000; 35:1111–29

Futures Trading with Transaction Costs

Karel Lavička [1]

Abstract. We specify a continuous time model of futures contract market where the price follows an arithmetic Brownian motion. In this market an agent is not allowed to have a short position in the contract or borrow money to make investment, but he or she can invest arbitrary portion of his or her wealth. With each transaction proportional transaction cost is paid and agent's goal is to maximize long-term logarithmic utility of the remaining wealth.

State of the system is described by a stochastic differential equation (SDE). Using Itô's formula we show that the information about the portion of invested wealth is sufficient for making optimal decisions.

We use Markov decision process with finite state space and discrete time to approximate SDE for the portion of invested wealth. The value function for approximating finite Markov chain is found by policy iteration, also known as Howard's algorithm. The optimal policy is determined by an interval which the portion of invested wealth has to be kept in. At the end we provide numerical results.

Keywords: Markov decision processes, policy iteration, Itô's formula, futures trading, transaction costs.

JEL classification: C44
AMS classification: 90C40

1 Introduction

The main purpose of this contribution is to show how an investor in the futures market may optimize his or her trading strategy. We begin with the description of the futures market model in the next section. The model is specified in continuous time and space. The main advantage is that in this setting we can easily use Itô's calculus to reduce the dimension of our problem.

In the third section we give a brief introduction into dynamic decision making in discrete time and space. The main part of this section is the description of Howard's policy iteration algorithm that finds the optimal solution of a given problem.

The fourth and fifth sections are about discrete approximation of the original agent's continuous problem. We specify here in detail a controlled Markov chain that is suitable for policy iteration and approximate controlled continuous time process in a sense of weak convergence. Results of a numerical procedure are shown for different initial parameters of the futures market.

In the conclusion we point out few remarks about the assumptions in our problem and the interpretation of the numerical results.

2 Trading in the futures market

In this section we describe in detail the mathematical model of investment in futures contract. We model it's price as an arithmetic Brownian motion, see (Janeček, Shreve [3]) or (Dostál [1]) for a short discussion of this choice. These papers are also the main references for our market model and reader can find more details in them. Our futures price is

$$F_t \;=\; F_0 + \mu t + \sigma W_t \quad \text{for} \quad t \geq 0,$$

[1]Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 186 75 Praha 8, Czech Republic, lavickak@karlin.mff.cuni.cz

where initial price F_0 and drift μ are arbitrary real numbers, volatility σ is a positive real number and W_t is a standard $\mathcal{F}_t$-adapted Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. In each time $t \geq 0$ the agent can buy or sell some portion of the futures contracts. Let us denote by $\mathcal{N}_t^+$ the number of futures contracts bought until time t. We assume that $\mathcal{N}_t^+$ is $\mathcal{F}_t$-adapted to exclude the possibility that the agent adapts his or her strategy to the uncertain price in the future. The decisions are based only on the information available until time t. Similarly $\mathcal{N}_t^-$ is a $\mathcal{F}_t$-adapted process that denotes the number of sold futures contracts. With each transaction the agent pays proportional transaction costs $\lambda_1 \geq 0$ for buying, $\lambda_2 \geq 0$ for selling respectively. Finally denoting agent's wealth process by $\mathcal{W}_t$ we can rigorously describe our model with stochastic differentials as follows

$$\begin{aligned} \mathrm{d}\mathcal{N}_t &= \mathrm{d}\mathcal{N}_t^+ - \mathrm{d}\mathcal{N}_t^-, \\ \mathrm{d}\mathcal{W}_t &= \mathcal{N}_t \mathrm{d}F_t - \lambda_1 \mathrm{d}\mathcal{N}_t^+ - \lambda_2 \mathrm{d}\mathcal{N}_t^-. \end{aligned} \tag{1}$$

We require the wealth $\mathcal{W}_t$ to be positive for all $t \geq 0$, thus $\mathcal{N}_t^+, \mathcal{N}_t^-$ are admissible transactions if and only if the agent always has enough money to cancel all the position in the market, i.e. $(\mathcal{W}_t, \mathcal{N}_t) \in \left\{(x,y) \in \mathbb{R}^2, x + \lambda_1 y > 0, x - \lambda_2 y > 0\right\}$ with probability one. In the space of all admissible transaction processes the agent wants to find a pair $\hat{\mathcal{N}}_t^+, \hat{\mathcal{N}}_t^-$ that provides maximal long-term expected logarithmic utility and also its value $\hat{\nu}$, that is for all

$$\limsup_{t\to\infty} \frac{1}{t} \mathbb{E} \log \mathcal{W}_t \;\leq\; \limsup_{t\to\infty} \frac{1}{t} \mathbb{E} \log \hat{\mathcal{W}}_t \;=\; \hat{\nu}, \tag{2}$$

where $\hat{\mathcal{W}}_t$ is the solution of (1) with $\mathcal{N}_t^+, \mathcal{N}_t^-$ replaced by $\hat{\mathcal{N}}_t^+, \hat{\mathcal{N}}_t^-$, i.e. $\hat{\mathcal{W}}_t$ is the optimal wealth proces.

In order to find the optimal transaction processes $\hat{\mathcal{N}}_t^+, \hat{\mathcal{N}}_t^-$, we define a portion of invested wealth as $\mathcal{R}_t = \mathcal{N}_t / \mathcal{W}_t$. We will show that one-dimensional and almost surely bounded controlled process $\mathcal{R}_t$ can fully describe the stochastic behaviour of the logarithmic utility that is to be maximized. The Itô's formula, see (Shreve [7]) Theorem 3.6, gives us

$$\begin{aligned} \mathrm{d}\mathcal{R}_t &= \mathcal{N}_t \mathrm{d}\mathcal{W}_t^{-1} + \mathcal{W}_t^{-1} \mathrm{d}\mathcal{N}_t + \mathrm{d}\left\langle \mathcal{N}, \mathcal{W}^{-1} \right\rangle_t \\ &= B(\mathcal{R}_t)\mathrm{d}t + S(\mathcal{R}_t)\mathrm{d}W_t + \mathrm{d}\mathcal{R}_t^+ - \mathrm{d}\mathcal{R}_t^-, \end{aligned} \tag{3}$$

where $\left\langle \mathcal{N}, \mathcal{W}^{-1} \right\rangle_t$ denotes the covariation process, see again (Shreve [7]), further

$$B(x) = x^2\sigma^2(x - \theta), \quad S(x) = -\sigma x^2, \quad \theta = \mu\sigma^{-2} \tag{4}$$

and

$$\mathrm{d}\mathcal{R}_t^+ = \mathcal{W}_t^{-1}(1 + \lambda_1 \mathcal{R}_t)\mathrm{d}\mathcal{N}_t^+, \quad \mathrm{d}\mathcal{R}_t^- = \mathcal{W}_t^{-1}(1 - \lambda_2 \mathcal{R}_t)\mathrm{d}\mathcal{N}_t^- \tag{5}$$

are changes in the invested portion that are caused by transactions, not just by price fluctuations. Note that if $\mathcal{N}_t$ is any admissible transaction process, then $\mathcal{R}_t \in (-\lambda_1^{-1}, \lambda_2^{-1})$ with probability one and from (5) we can now see that $\mathcal{R}_t^+$ and $\mathcal{N}_t^+$ generate equivalent Lebesgue-Stieltjes measures on the real line, thus using Radon-Nikodym theorem we have also

$$\mathrm{d}\mathcal{N}_t^+ = \mathcal{W}_t(1 + \lambda_1 \mathcal{R}_t)^{-1}\mathrm{d}\mathcal{R}_t^+, \quad \mathrm{d}\mathcal{N}_t^- = \mathcal{W}_t(1 - \lambda_2 \mathcal{R}_t)^{-1}\mathrm{d}\mathcal{R}_t^-, \tag{6}$$

which describes the size of transaction in terms of change in the invested portion. By combining (1) with (6) and the definition of $\mathcal{R}_t$, one gets

$$\mathrm{d}\mathcal{W}_t = \mathcal{R}_t\mathcal{W}_t(\mu\mathrm{d}t + \sigma\mathrm{d}W_t) - \mathcal{W}_t\lambda_1(1 + \lambda_1\mathcal{R}_t)^{-1}\mathrm{d}\mathcal{R}_t^+ - \mathcal{W}_t\lambda_2(1 - \lambda_2\mathcal{R}_t)^{-1}\mathrm{d}\mathcal{R}_t^-. \tag{7}$$

The Itô's formula used again gives

$$\mathrm{d}\log\mathcal{W}_t = \frac{1}{2}\sigma^2(2\theta\mathcal{R}_t - \mathcal{R}_t^2)\mathrm{d}t + \sigma\mathcal{R}_t\mathrm{d}W_t - \lambda_1(1 + \lambda_1\mathcal{R}_t)^{-1}\mathrm{d}\mathcal{R}_t^+ - \lambda_2(1 - \lambda_2\mathcal{R}_t)^{-1}\mathrm{d}\mathcal{R}_t^- \tag{8}$$

and the expectation in (2) now can be written as

$$\mathbb{E}\log\mathcal{W}_t = \mathbb{E}\int_0^t \frac{1}{2}\sigma^2(2\theta\mathcal{R}_s - \mathcal{R}_s^2)\mathrm{d}s - \mathbb{E}\int_0^t \lambda_1(1 + \lambda_1\mathcal{R}_s)^{-1}\mathrm{d}\mathcal{R}_s^+ - \mathbb{E}\int_0^t \lambda_2(1 - \lambda_2\mathcal{R}_s)^{-1}\mathrm{d}\mathcal{R}_s^-, \tag{9}$$

because the term $\sigma\mathcal{R}_s$ is bounded with probability one, hence $\int_0^t \sigma\mathcal{R}_s\mathrm{d}W_s$ is a martingale with zero expectation. Formula (9) allows us to reformulate the original goal into equivalent one. The agent now wants to find a pair $\hat{\mathcal{R}}_t^+$, $\hat{\mathcal{R}}_t^-$ of $\mathcal{F}_t$-adapted processes that control (as in (3)) the portion of invested wealth $\mathcal{R}_t$ in optimal way, i.e. (2) holds, where $\hat{\mathcal{W}}_t$ is a solution of (7) with $\mathcal{R}_t^+$, $\mathcal{R}_t^-$, $\mathcal{R}_t$ replaced by $\hat{\mathcal{R}}_t^+$, $\hat{\mathcal{R}}_t^-$, $\hat{\mathcal{R}}_t$. The optimal transaction processes $\hat{\mathcal{N}}_t^+, \hat{\mathcal{N}}_t^-$ are connected with $\hat{\mathcal{R}}_t^+$, $\hat{\mathcal{R}}_t^-$ via equations (6).

3 Markov decision processes - policy iteration

In this section we specify a model for dynamic decision making in discrete time. The main building block is a finite valued controlled Markov chain with transition probabilities $P(u) = \{p_{ij}(u_i)\}_{i,j\in S}$, where S is a finite set of states and $u = \{u_i \in \mathcal{U}_i, i \in S\} \in \mathcal{U}$ is a vector of admissible decisions. We assume that for each addmisible decision vector $u \in \mathcal{U}$ there exists a homogeneous ergodic Markov chain X^u with transition probabilities $P(u)$ and unique invariant distribution $\pi(u)$. Quality of $u \in \mathcal{U}$ is measured by a reward function Z that assigns a real value $Z(i, j, u_i)$ to each transition from the state i to the state j with the decision u_i selected. Decision maker's goal is to maximize long-term expected gain, i.e. he or she wants to find a decision $\hat{u} \in \mathcal{U}$ and a real value $\hat{\nu}$ such that for all $u \in \mathcal{U}$

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\left(\sum_{k=1}^{n} Z(X_k^u, X_{k+1}^u, u_{X_k^u})\right) \;\le\; \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\left(\sum_{k=1}^{n} Z(X_k^{\hat{u}}, X_{k+1}^{\hat{u}}, \hat{u}_{X_k^{\hat{u}}})\right) \;=\; \hat{\nu}. \tag{10}$$

For a given decision $u \in \mathcal{U}$ and its long-term value $\nu(u)$, we may deduce, see theorem 1.8.3 in (Norris [5]),

$$\begin{aligned} \nu(u) \;&=\; \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\left(\sum_{k=1}^{n} Z(X_k^u, X_{k+1}^u, u_{X_k^u})\right) \;=\; \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}\sum_{i,j\in S} \mathbb{E} I_{\{X_k^u=i, X_{k+1}^u=j\}} Z(i,j,u_i) \\ &=\; \sum_{i,j\in S} Z(i,j,u_i)p_{ij}(u_i) \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} p_i^{(k)}(u) \;=\; \sum_{i,j\in S} Z(i,j,u_i)p_{ij}(u_i)\pi_i(u), \end{aligned} \tag{11}$$

where $p_i^{(k)}(u) = \mathbb{P}(X_k^u = i)$.

To solve the optimization problem $\max_{u\in\mathcal{U}} \nu(u)$ one can use policy iteration, also known as Howard's algorithm, see (Howard [2]), which is a procedure that seeks for the optimal decision directly in the space of decisions, rather than indirectly via value functions. General theory for policy iteration can be found in (Puterman [6]). We show here a brief introduction. Let us denote by 1 the unit vector and by $C(u)$ the vector of expected rewards, i.e. $C_i(u_i) = \sum_{j\in S} Z(i,j,u_i)p_{ij}(u_i)$ for each $i \in S$. From (11) can be seen that $\nu(u) = \pi'(u)C(u)$, where $\pi'(u)$ denotes the transposition of the invariant distribution.

Theorem 1 (Policy iteration algorithm)**.** *Let $\mathcal{U}$ be finite, $u^0 \in \mathcal{U}$ be an arbitrary decision vector and for $n \in \mathbb{N}$ be $u^n \in \mathcal{U}$ such that*

$$u_i^n \;\in\; \operatorname*{argmax}_{u_i\in\mathcal{U}_i} \left\{\sum_{j\in S} p_{ij}(u_i)W_j(u^{n-1}) + C_i(u_i)\right\}, \quad i \in S, \tag{12}$$

where $(W(u^{n-1}), \nu(u^{n-1}))$ is a solution of

$$W(u^{n-1}) + 1\nu(u^{n-1}) \;=\; P(u^{n-1})W(u^{n-1}) + C(u^{n-1}). \tag{13}$$

Then $\lim_{n\to\infty} \nu(u^n) = \hat{\nu}$, where $\hat{\nu}$ is the optimal value defined in (10)*.*

Proof. From expression (12) we can see for $u \in \mathcal{U}$

$$P(u^n)W(u^{n-1}) + C(u^n) \ge P(u)W(u^{n-1}) + C(u), \tag{14}$$

what for special choice $u = u^{n-1}$ after multiplying both sides by $\pi'(u^n)$ gives

$$\begin{aligned} \nu(u^n) \;=\; \pi'(u^n)C(u^n) \;&\ge \pi'(u^n)\Big(P(u^{n-1})W(u^{n-1}) + C(u^{n-1}) - P(u^n)W(u^{n-1})\Big) \\ &= \pi'(u^n)\Big(P(u^{n-1})W(u^{n-1}) + C(u^{n-1}) - W(u^{n-1})\Big) \\ &= \pi'(u^n)1\nu(u^{n-1}) \\ &= \nu(u^{n-1}). \end{aligned}$$

Under ergodicity assumption we have $\pi_i(u) > 0$ for all $i \in S$ and $u \in \mathcal{U}$, thus $\nu(u^n) > \nu(u^{n-1})$ if and only if at least one inequality in (14) is strict. Sequence $\nu(u^n)$ is monotone and $\mathcal{U}$ is finite, therefore there exists $n_0 \in \mathbb{N}$ such that $\nu(u^{n_0}) = \nu(u^{n_0-1})$. This together with (13) implies

$$P(u^{n_0})W(u^{n_0-1}) + C(u^{n_0}) \;=\; P(u^{n_0-1})W(u^{n_0-1}) + C(u^{n_0-1}) \;=\; W(u^{n_0-1}) + 1\nu(u^{n_0-1}) \tag{15}$$

Now using (14) and (15) together we have for all $u \in \mathcal{U}$

$$W(u^{n_0-1}) + 1\nu(u^{n_0-1}) \;\geq\; P(u)W(u^{n_0-1}) + C(u). \tag{16}$$

The final result $\nu(u^{n_0-1}) \geq \nu(u)$ can be obtained from (16) by multiplying both sides by $\pi'(u)$. □

4 Discrete approximation of the problem

Numerical solution of the problem (2) with conditions (3) and (9) relies on the probabilistic methods that approximate random process $\mathcal{R}_t$ with Markov decision process, where its decisions correspond to $\mathcal{R}_t^+$ and $\mathcal{R}_t^-$. After solving this finite problem problem with policy iteration algorithm, we will show in the next section how to interpret the result in the original continuous problem (2). A costruction of the approximate process is natural, however its convergence proof is nontrivial and it is not discussed here, but it can be seen in (Kushner and Dupuis [4]).

Let $h > 0$ is a small aproximation parameter and $S_h = \{kh, k \in \mathbb{Z}\} \cap (-\lambda_1^{-1}, \lambda_2^{-1})$ be a state space for a Markov chain X^u that in our discrete setting corresponds to the process $\mathcal{R}_t$. We will use the dynamics in (3) to deduce the transition probabilities of X^u. They depend on the controls $\mathcal{R}_t^+$ and $\mathcal{R}_t^-$, thus $\mathcal{U}$ needs to be specified at first. We use symbol $\nearrow$ for buying the futures contract, $\searrow$ for selling it and $\rightarrow$ for no trading. The possibility of buying and selling in the same moment is excluded from our considerations, because it obviously brings higher transaction costs than no trading with the same result. The set of admissible decisions $\mathcal{U}$ is given by the following rules.

$$\begin{aligned}
\overline{ih}_1 &= \max S_h, & \mathcal{U}_{\overline{ih}_1} &= \{\searrow\},\\
\overline{ih}_2 &= \max S_h \setminus \overline{ih}_1, & \mathcal{U}_{\overline{ih}_2} &= \{\rightarrow, \searrow\},\\
\underline{ih}_1 &= \min S_h, & \mathcal{U}_{\underline{ih}_1} &= \{\nearrow\},\\
\underline{ih}_2 &= \min S_h \setminus \underline{ih}_1, & \mathcal{U}_{\underline{ih}_2} &= \{\nearrow, \rightarrow\},\\
i \in S_h \setminus \{\overline{ih}_1, \overline{ih}_2, \underline{ih}_1, \underline{ih}_2\}, &\ \text{then} & \mathcal{U}_{ih} &= \{\searrow, \rightarrow, \nearrow\}.
\end{aligned}$$

In the state $\overline{ih}_1$, i.e. close to the upper bound of the feasible region $(-\lambda_1^{-1}, \lambda_2^{-1})$, the agent is forced to sell the contract to keep his or her wealth positive. Similar reasoning gives the constraints for $\overline{ih}_2$, $\underline{ih}_1$ and $\underline{ih}_2$. At first we specify the dynamics of X^u if the agent does not buy or sell any contract, i.e. $u_{ih} = \rightarrow$. Let the functions B and S be given by (4) and set $\Delta t = h^2 / \max_{x \in (-\lambda_1^{-1}, \lambda_2^{-1})} S(x)$ and $\Delta X_n^u = X_{n+1}^u - X_n^u$. According to (Kushner and Dupuis [4]), one needs the same expectation and variance of increments for X^u and $\mathcal{R}_t$, that is

$$\mathbb{E}_{ih}\Delta X_1^u = B(ih)\Delta t + o(\Delta t), \quad \mathbb{E}_{ih}(\Delta X_1^u - \mathbb{E}_{ih}\Delta X_1^u)^2 = S^2(ih)\Delta_t + o(\Delta t) \tag{17}$$

is required, where $\mathbb{E}_{ih}(\cdot) = \mathbb{E}(\cdot | X_1^u = ih)$. This can be satisfied by setting

$$\begin{aligned}
p_{ih,(i+1)h}(\rightarrow) &= \frac{S^2(ih)\Delta t + hB(ih)\Delta t}{2h^2},\\
p_{ih,(i-1)h}(\rightarrow) &= \frac{S^2(ih)\Delta t - hB(ih)\Delta t}{2h^2},\\
p_{ih,ih}(\rightarrow) &= 1 - \frac{S^2(ih)\Delta t}{h^2}
\end{aligned} \tag{18}$$

for h sufficiently small such that all transition probabilities in (18) are between zero and one. From (3) can be seen that the controls $\mathcal{R}_t^+$ and $\mathcal{R}_t^-$ can move the process $\mathcal{R}_t$ immediately up or down. The same should hold for X^u under decisions $\nearrow$ and $\searrow$. A small parameter $\epsilon > 0$ is introduced not to break the ireducibility of X^u that is needed for policy iteration. But for $\epsilon \to 0_+$ the local consistency of X^u in the sense of (17) is almost satisfied if we set

$$\begin{aligned}
p_{ih,(i+2)h}(\nearrow) &= (1-\epsilon)p_{(i+1)h,(i+2)h}(\rightarrow),\\
p_{ih,(i+1)h}(\nearrow) &= \epsilon p_{ih,(i+1)h}(\rightarrow) + (1-\epsilon)p_{(i+1)h,(i+1)h}(\rightarrow),\\
p_{ih,ih}(\nearrow) &= \epsilon p_{ih,ih}(\rightarrow) + (1-\epsilon)p_{(i+1)h,ih}(\rightarrow),\\
p_{ih,(i-1)h}(\nearrow) &= \epsilon p_{ih,(i-1)h}(\rightarrow),
\end{aligned} \tag{19}$$

what corresponds to the immediate move to the state $(i+1)h$ with probability $1-\epsilon$ and then the behaviour of X^u is the same as in (18). Similarly

$$\begin{aligned} p_{ih,(i+1)h}(\searrow) &= \epsilon p_{ih,(i+1)h}(\rightarrow), \\ p_{ih,ih}(\searrow) &= \epsilon p_{ih,ih}(\rightarrow) + (1-\epsilon) p_{(i-1)h,ih}(\rightarrow), \\ p_{ih,(i-1)h}(\searrow) &= \epsilon p_{ih,(i-1)h}(\rightarrow) + (1-\epsilon) p_{(i-1)h,(i-1)h}(\rightarrow), \\ p_{ih,(i-2)h}(\searrow) &= (1-\epsilon) p_{(i-1)h,(i-2)h}(\rightarrow). \end{aligned} \tag{20}$$

The remaining part that needs to be specified for the approximating Markov decision process is the reward function Z. From comparison of (2), (10) and (9) can be seen that

$$\begin{aligned} Z(ih, jh, \rightarrow) &= \frac{1}{2}\sigma^2(2\theta ih - (ih)^2)\Delta t, \\ Z(ih, jh, \nearrow) &= Z(ih, jh, \rightarrow) - h\lambda_1(1+\lambda_1 ih)^{-1}, \\ Z(ih, jh, \searrow) &= Z(ih, jh, \rightarrow) - h\lambda_2(1-\lambda_2 ih)^{-1} \end{aligned}$$

is an appropriate choice for the reward function.

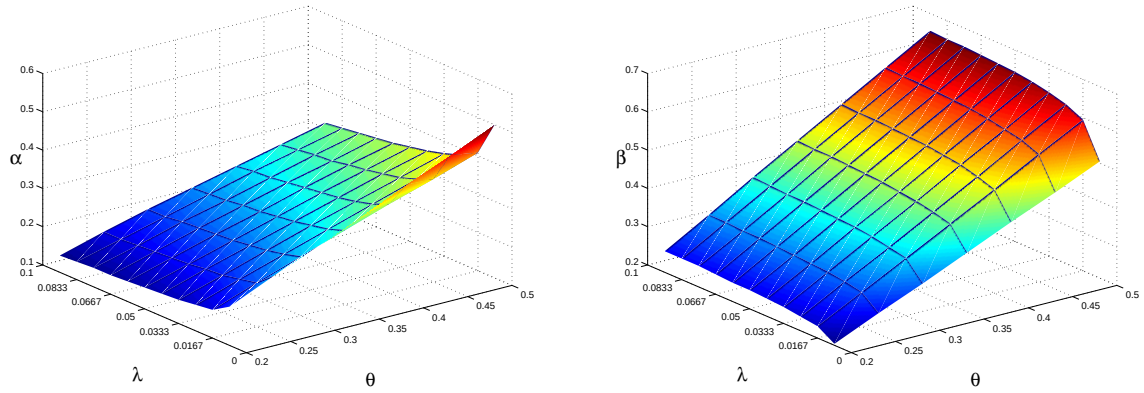
Next section shows optimal decisions that were obtained by policy iteration algorithm used for X^u specified here.

5 Numerical results

We run policy iteration algorithm with parameters h such that $|S_h| = 300$ and $\sigma = 1$ and variable parameters $\lambda = \lambda_1 = \lambda_2$ and θ. A numerical procedure showed that the optimal decision vector $\hat{u}$ has the following form.

$$\hat{u}_{ih} = \begin{cases} \nearrow & ih \leq \alpha \\ \rightarrow & \alpha < ih < \beta \\ \searrow & \beta \leq ih \end{cases}$$

Parameters α, β and $\hat{\nu}$ that describe the optimal solution are shown in the figures 1 and 2 for different transaction costs and drift parameters.



(a) Optimal lower bound α of the no trading region (b) Optimal upper bound β of the no trading region

Figure 1: Optimal no trading region

6 Conclusion

We showed in this contribution how the invesor may optimize his or her long-term investment in the futures contract market. The most important and problematic assumption that we made was about

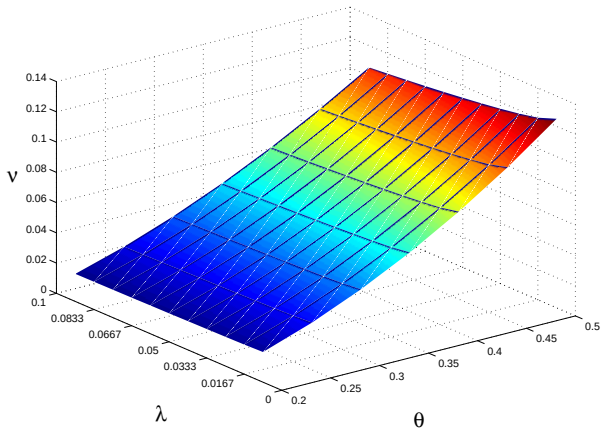


Figure 2: Optimal long-term utility value ν

futures price behaviour. We describe it by the arithmetic Brownian motion with constant and known parameters. Stochastic drift and volatility parameters would generalize our model. Nevertheless there is no reason for different qualitative properties of our result in the generalized setting.

With reasonable values of drift and transaction costs it turned out, see figure 2, that an increase of the drift parameter has significantly higher impact on the long-term utility than a decrease in the transaction costs.

We also showed, see figure 1, that the optimal trading strategy has to keep the ratio of invested wealth within no trading interval in order to reduce rebalancing of the investment which brings high transaction costs. Position of the interval depends on the drift parameter and its width depends on the size of transaction costs.

Acknowledgements

This work was supported by the grant No. P402/12/6097 of the Czech Science Foundation. I would like to also thank to SVV-2014-260105 for partial support.

References

[1] Dostál, P.: Futures Trading with Transaction Costs, *Proceedings of ALGORITMY* (2009), 419–428.

[2] Howard, R.: Dynamic Programming and Markov Processes, *The M.I.T. Press* (1960).

[3] Janeček, K., Shreve, S.: Futures Trading with Transaction Costs, *Illinois Journal of Mathematics* (2010), 1239–1284. 419–428.

[4] Kushner, H. J., Dupuis, P.: *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer, 2001.

[5] Norris, J. R.: *Markov Chains*. Cambridge University Press, 1998.

[6] Puterman, M. L.: *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, 2005.

[7] Shreve, S.: *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, 2008.

A Comprehensive Cash-Flow Valuation Approach Using Real Options

Yuri Lawryshyn[1], Michael Vinelli[2]

Abstract. Valuation techniques utilized by investment banks have grown in sophistication and complexity. Discounted Cash Flow (DCF) analysis is the most dominant and widely accepted technique for valuing a broad range of projects and companies. However, the DCF analysis is typically very sensitive to the discount rate, leading to a broad range of valuation estimates. Furthermore, especially for the valuation of early start-ups, the methodology does not properly account for potential company termination. A real options technique called the "Matching Cash-Flows" (MCF) method was recently developed to value early stage investment projects. The major advantage of the method is that it does not require a discount rate estimate, but rather, relies on the distribution of cash-flow projections to quantify the riskiness of the project or company. Furthermore, the method can properly account for bankruptcy. In this paper, we show how the MCF method can be used effectively, in place of DCF analysis. Two case studies are presented to show the advantage of the methodology.

Keywords: Real options, valuation, discounted cash-flow

JEL Classification: C61
AMS Classification: 90C15

1 Introduction

In Investment Banking, research analysts are often tasked with performing valuations on a wide variety of companies from different industries. Valuations predominately take the form of Discounted Cash Flow (DCF) analysis and are often supplemented with other ratio analysis techniques. Widely regarded as the most proficient valuation method, DCF analysis has limitations as pointed out by many authors and scholars. Real Option Analysis (ROA) is a valuation technique that addresses and provides a solution to the limitations of DCF analysis. It provides a means to deal with cash-flow uncertainties and managerial flexibility. ROA attempts to quantify the inherent investment flexibility within real-world investment opportunities. A popular example of this flexibility is found in pharmaceutical companies where the development of new drugs are done in increasingly expensive stages and managers can terminate the projects when it becomes clear that there is a low probability of success in the development. Thus, there exist inherent call-like options in the projects. In academic literature, ROA has been recognized as a superior method for the valuation of early stage investments compared to standard evaluation techniques such as net present value and discounted cash-flow analysis [1]. While, ROA is especially well suited for the valuation of start-up companies and R&D projects; its applicability extends beyond this scope.

Despite ROA's prevalence in academic literature (e.g. [2] and [1]), its adoption in industry is low. A survey of the Fortune 1000 largest companies found that only 14.3% of respondents use some type of ROA as part of their capital budgeting process [3]. More recently [4] found that only 8.1% of large Canadian firms surveyed used ROA when making decisions on investment projects. Industry's reluctance to more widely adopt ROA can be attributed to leading approaches that lack financial rigor or are not practically implementable. In response to the issues facing ROA, [5] have developed a novel ROA valuation technique called the Matching Cash-Flows (MCF) method that is both practical to implement and consistent with financial theory. The MCF method is built upon pessimistic, average and optimistic cash flow estimates that, as the authors suggest, managers are already likely to provide. The method requires minimal subjectivity with respect to parameter estimation and properly accounts for both market and private risks. Furthermore, it ensures that cash flows are appropriately correlated across time periods. The gap that remains to be filled is the practical implementation of an ROA method and a comparison of ROA valuation on historical companies/projects to valuation techniques used at the time. This paper attempts to display the merits of the MCF method using case studies comparing the MCF method to standard DCF techniques.

[1] University of Toronto, Centre for Management of Technology and Entrepreneurship (CMTE), 200 College St., Toronto, Ontario, Canada, yuri.lawryshyn@utoronto.ca.
[2] University of Toronto, CMTE, 200 College St., Toronto, Ontario, Canada, michael.vinelli@mail.utoronto.ca.

2 Methodology

The MCF method framework mirrors the derivation of Black-Scholes [6] pricing, as discussed in great depth in [5]. Here we provide a brief overview of the methodology. We begin by assuming that managers have supplied low, medium and high sales and gross margin percent (GM%) estimates from times $t = T_1, T_2, \dots, T_k$, namely, $S_1, S_2, \dots, S_k$, and $M_1, M_2, \dots, M_k$. Furthermore, the MCF method assumes there is a traded index whose price, I_t, follows geometric Brownian motion,

$$dI_t = \mu I_t dt + \sigma I_t dB_t \tag{1}$$

where μ and σ are the drift and volatility parameters and are assumed to be constant and B_t is a standard Brownian motion (BM) under the real-world measure P. Next, it is assumed that there exist both a *sales sector indicator*, X_t, and a *gross margin percent sector indicator*, Y_t,

$$dX_t = \rho_{SI} dB_t + \sqrt{1-\rho_{SI}^2} dW_t^S, \tag{2}$$

$$dY_t = \rho_{SM} dX_t + \sqrt{1-\rho_{SM}^2} dW_t^M, \tag{3}$$

where W_t^S and W_t^M are standard Brownian motions in the real-world measure, independent of each other and B_t, and ρ_{SI} and ρ_{SM} are correlations. The BMs X_t and Y_t are used to drive the sales and GM% estimates.

In the MCF method, a collection of functions φ_k^S and φ_k^M are introduced such that

$$S_k = \varphi_k^S(X_{T_k}) \text{ and } M_k = \varphi_k^S(Y_{T_k}) \tag{4}$$

and it can be shown that for a given managerial specified sales (or GM%) distribution $F^*(v)$ at time T,

$$\varphi^S(x) = F^{*-1}\left(\Phi\left(\frac{x - x_0}{\sqrt{T}}\right)\right). \tag{5}$$

The k-th cash-flow, i.e. the cash-flow at time T_k, can be calculated as $v_k = \varphi_k^S(\varphi_k^M - \kappa_k) - \alpha_k$, where κ_k is a deterministic variable cost and α_k is a fixed cost, both at time T_k.

Following [5], the risk-neutral process for I_t is given by

$$dI_t = rI_t dt + \sigma I_t d\hat{B}_t, \tag{6}$$

where r is the risk-free rate, assumed to be constant, and $\hat{B}_t$ is a Brownian motion under the risk-neutral measure, Q. Applying the minimum martingale measure, the processes for X_t and Y_t become

$$dX_t = \hat{\nu} dt + \rho_{SI} d\hat{B}_t + \sqrt{1-\rho_{SI}^2} d\widehat{W}_t^S, \tag{7}$$

$$dY_t = \hat{\gamma} dt + \rho_{SI}\rho_{SM} d\hat{B}_t + \rho_{SM}\sqrt{1-\rho_{SI}^2} d\widehat{W}_t^S + \sqrt{1-\rho_{SM}^2} d\widehat{W}_t^M, \tag{8}$$

where $\hat{B}_t$, $\widehat{W}_t^S$ and $\widehat{W}_t^M$ are standard uncorrelated Brownian motions under the risk-neutral measure Q and the risk-neutral drifts of the indicators are $\hat{\nu} = -\rho_{SI}\frac{\mu - r}{\sigma}$ and $\hat{\gamma} = -\rho_{SI}\rho_{SM}\frac{\mu - r}{\sigma}$.

Different to [5], here we assume that it may be optimal to exit the project (business) if the expected future value of the project is negative. In this case, our valuation takes on the form of an American option at $t > T_0$. At some time $T_0 < T_1$, the value of the cash-flows can be written as

$$V_{T_0}(S_{T_0}, M_{T_0}) = \sup_{\tau \in \mathcal{T}} \mathbb{E}^{\mathbb{Q}}\left[\sum_{k=1}^{n} e^{-r(T_k - T_0)} v_k(S_{T_k}, M_{T_k}) \middle| S_{T_0}, M_{T_0}\right], \tag{9}$$

where $\mathcal{T}$ represents a set of stopping times on the interval $[T_0, T_n]$ and $\mathbb{E}^{\mathbb{Q}}[\cdot]$ represents the expectation with respect to the risk-neutral measure $\mathbb{Q}$. In the real option setting, there may be a cost, K, that must be incurred at T_0 to attain the cash-flows. Furthermore, managers may specify an idiosyncratic but deterministic probability of success, p_0. Thus, the real option value at t, where $0 \le t < T_0$, is given by

$$RO_t(S, M) = p_0 e^{-r(T_0 - t)} \mathbb{E}^{\mathbb{Q}}\left[\left(V_{T_0}(S_{T_0}, M_{T_0}) - K\right)^+ \middle| S_t = S, M_t = M\right]. \tag{10}$$

Following [5], we define $G_t(X, Y) = RO_t(S(X), M(Y))$ and apply Ito's lemma and standard techniques to get

$$rG = \frac{\partial G}{\partial t} + \hat{\nu}\frac{\partial G}{\partial x} + \hat{\gamma}\frac{\partial G}{\partial y} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2} + \frac{1}{2}\frac{\partial^2 G}{\partial y^2} + \rho_{SM}\frac{\partial^2 G}{\partial x \partial y}. \tag{11}$$

Equation (11) is solved explicitly. The boundary condition at $t = T_n$ is given by $G(X_{T_n}, Y_{T_n}) = v_n$ and at each cash-flow date T_k, v_k is added to $G(X_{T_n}, Y_{T_n})$. For all $t > T_0$ we set $G(X_t, Y_t) = \max(G(X_t, Y_t), 0)$ and at the option strike data we set $G(X_{T_0}, Y_{T_0}) = \max(G(X_{T_0}, Y_{T_0}) - K, 0)$. Furthermore, it is easy to extend the numerical valuation to account for multiple options on options. We have implemented this approach in Microsoft Excel using Visual Basic for Applications (VBA) and the results presented in Cases I and II are computed using this tool.

3 Case Studies

This Section reviews two case studies in which the MCF method was applied to pre-existing company valuations. Our sponsoring investment bank has provided a DCF valuation of a biotech company as well as a flooring distributor. The estimated revenues, gross margin %, fixed and variable costs, net working capital requirements and other parameters from these valuations were translated to the MCF Tool to build comparisons between the MCF method and the standard DCF approach currently used by the investment bank.

3.1 Case I: Biotech Company

The biotech company spreadsheet that was provided to us contains a significant amount of information related to the sources of revenue and cost projections. Factors such as the number of patients the company has access to, the prevalence of the drug, potential market share, and others were used to develop revenue and cost projections. These calculations were required for both the DCF and MCF analyses. It was on the basis of how these numbers were used to formulate a value for the company that differentiates the two methodologies. Within the biotech company spreadsheet, there exit conservative and aggressive estimates for the revenues and costs in each year. The revenues begin after 4 years of investment during trials and testing. Each year, free cash flow (FCF) for the company was calculated as

$$FCF = Revenue - COGS - Costs - Taxes - \Delta Working\ Capital - Capital\ Expenitures \quad (12)$$

until certain patents expire. The free cash-flows, after these patents expire, were insignificant to the value of the company. There exist both conservative and aggressive estimates for the FCFs in each year in the data provided to us. The FCFs were discounted at different hurdle rates to determine their present value. An 80% risk adjustment was then multiplied with the net present value of the FCFs to arrive at the final valuation for the biotech company. The risk adjustment represents the probability the company will pass Phase II FDA approval with their drug and realize any revenues beyond the approval. The conservative-estimate and aggressive-estimate valuations for the biotech company are summarized in Table 1 for different discount rates. The discount rate the bank had chosen is highlighted (i.e. 12.5%).

Table 1. Biotech company valuations at different discount rates

Discount Rate	Conservative Estimate	Aggressive Estimate
10%	$ 578,789	$ 1,028,961
12.5%	$ 471,245	$ 836,995
15%	$ 386,216	$ 685,590
17.5%	$ 318,419	$ 565,182
20%	$ 263,938	$ 468,674

The MCF analysis used the same cash flow estimates as the DCF, except that they needed to be translated to work with the MCF tool. First, a median estimate for the cash flows was determined using the average of the high and low estimates. Second, the 80% risk adjustment was translated. After discussions with the analysts, it was determined that this number was arbitrarily picked and somewhat subjective. Since there are 4 years of investment in the company before any revenues are realized, we assumed a probability of success of 94% for each year ($0.8^{1/4} \approx 0.94$) (Table 2). The net working capital and capital expenditures exist as a percentage of revenues in the biotech company spreadsheet. This matches with the cash flow estimate requirements of the MCF Tool.

Table 2. Investments and probability of success information for the biotech company

Strike Price (K)	Time of Strike (y)	Probability of Success
$5000	1	94%
$10,875	2	94%
$11,500	3	94%
$28,714	4	94%

One of the significant advantages of the MCF method is that it requires minimal subjective parameterization. Aside from estimates associated with the cash-flows themselves, the MCF method requires a few market parameters, which we assume are easily attainable, as well as the correlations (ρ_{SI}, ρ_{SM}) and the probability of success of passing through the required stages. The market parameters were assumed as follows: 5% risk-free rate, 16% traded index volatility and 11% expected growth for the traded index. The correlations were assumed as follows: $\rho_{SI} = 50\%$, $\rho_{SM} = 0$. The tax rate was given as 25%.

While the risk-free rate and tax rate were determined based on the information given in the biotech company spreadsheet, the others are estimated separately. The traded index volatility and expected growth were chosen based on the statistics for 'XLV', which is an ETF that tracks the Health Care Sector of the S&P 500 [7]. The revenue/GM% correlation was estimated to be zero because the biotech company spreadsheet used a constant gross margin regardless of the conservative or aggressive revenue estimates. The revenue/traded index correlation was estimated and a sensitivity analysis on the parameter follows. Table 3 displays the value of the biotech company using the MCF method. It also shows how the valuation changes as different company specific parameters are varied.

Table 3. Biotech company valuations using the MCF method Tool

Variable	Value
Baseline	$ 865,421
Probability of success (90-90-90-90) years 1-4	$ 726,133
0% Revenue/traded index correlation	$ 931,179
30% Revenue/traded index correlation	$ 891,396
80% Revenue/traded index correlation	$ 828,074
100% Revenue/traded index correlation	$ 804,678
20% Revenue/GM% correlation	$ 864,498

The valuation results from Table 1 and Table 3 are directly comparable. Both the DCF and MCF methods use the same cash flow estimates and differ only in terms of a few estimated parameters. The DCF applies a discount rate and risk adjustment to the cash flow estimates while the MCF method applies a risk-free rate, information about a related traded index, and two correlations. Although one cannot argue which valuation is more valid, comments on parameter estimation and sensitivity to these parameters can be made. The valuations show that the MCF method value matches more closely with the aggressive DCF analysis. This result is expected since the value of the real option inherent in the company has been accounted for by the MCF analysis.

The MCF method is less sensitive to its parameters (as evidenced in Table 3) than the DCF model (Table 1). Additionally, the parameters estimated for the MCF method are less subjective than the DCF parameters. First, the MCF method uses information about a related index and supposed correlations while the DCF has a rather subjective discount rate. Second, the methods differ in their perception of the risk associated with the Phase III FDA approval. The MCF method applies real options on investment dates, which is understood in this context as applying a probability of success through the trials. On the other hand, the DCF applies a subjective risk adjustment to the NPV of the company that is theoretically inaccurate. In fact, the DCF method does not properly account for the fact that managers have the option to stop investing if, at some point in the future, it is warranted to do so, and this is borne out of the real options analysis with the MCF method.

3.2 Case II: Flooring Distribution

Case II seeks to demonstrate the merits of the MCF method with a direct comparison to a simplified DCF. The Case II DCF analysis is a valuation of a flooring distributor with 5 years' worth of cash flow estimates and an estimated terminal growth rate beyond that. It does not contain the typical features that would suggest using an ROA approach (such as the Phase III trial results found in Case I). This case allows for a direct comparison between typical parameter estimation in a DCF and the parameters estimated in the MCF method's. Each year, a free cash flow (FCF) for the company was calculated as

$$FCF = (Revenue \times EBITDA\ Margin\%) - Taxes - \Delta Working\ Capital - Capital\ Expenitures \quad (13)$$

for five years. In the sixth year a terminal value was calculated using; the fifth year FCF, a terminal growth rate and the Gordon Growth Model. In the data provided to us, there exist high, medium and low estimates of revenue and EBITDA Margin % in each year that produces three different valuations for the company. A weighted average cost of capital (WACC) discount rate was used. The return on equity was derived using the capital asset pricing model (CAPM), however, additional terms were added. The investment bank analysts add a 'small company premium' and 'company specific risk premium', which can be simplified to represent liquidity risk. The parameters for the valuation are summarized in Table 4. Using these assumptions and the estimates for revenue and EBITDA Margin %, Table 5 displays the valuation of the flooring distributor discounted at the WACC.

Another important aspect of this valuation is its sensitivity to the parameter estimates, specifically the discount rate and terminal growth rate. Table 6 shows a sensitivity analysis for the median case. As evidenced in Table 6, the DCF valuation is somewhat sensitive to the discount rate or the terminal growth rate. The variability in the DCF value of the flooring distributor lies especially in the difference between the worst-case and best-case scenarios.

Table 4. Assumptions used to derive the DCF value of the flooring distributor.

Item	Value
Tax Rate	33%
Discount Rate (WACC)	13.5%
Liquidity risk premium (Built into WACC)	7%
Terminal growth rate	2%

Table 5. Flooring distributor DCF valuation discounted at the WACC

Case	Company Value Estimate ($000s)
Low	$36,735
Medium	$46,918
High	$54,564

Table 6. Median flooring distributor DCF with varying WACC and terminal growth ($000s)

	Terminal Value Growth Rate				
WACC	1.0%	1.5%	2.0%	2.5%	3.0%
12.0%	50,483	52,027	53,725	55,602	57,688
12.5%	48,340	49,724	51,241	52,909	54,753
13.0%	46,374	47,621	48,982	50,471	52,110
13.5%	44,565	45,692	46,918	48,254	49,718
14.0%	42,895	43,917	45,025	46,228	47,542
14.5%	41,347	42,277	43,282	44,370	45,553
15.0%	39,910	40,759	41,673	42,660	43,729

In the MCF analysis we used the same cash-flow estimates as the DCF, except the terminal growth was treated differently. Specifically, the MCF method performs the valuation from the last cash flow back to the first, through dynamic programming – no optionality was built in, except to account for the fact that a negative future expected value of cash-flows was set to zero to account for the stopping time, consistent with equation (9). A model that forecasts 'forever' into the future as the Gordon Growth Model does will not apply to the MCF method. Instead estimates 15 years beyond the final cash flow were included in the tool. This 15-year forecasting accounts for nearly 90% of the Gordon Growth Model terminal value. This was deemed sufficient. As discussed above for Case I, to complete the MCF method valuation, the MCF model parameters must be chosen. The market parameters were assumed as follows: 5% risk-free rate, 20% traded index volatility and 12% expected growth for the traded index. The correlations were assumed as follows: $\rho_{SI} = 50\%$, $\rho_{SM} = 5\%$. The tax rate was given as 33%. While the risk-free rate and tax rate can be surmised from the flooring distributor spreadsheet, the others are estimated separately. The traded index volatility and expected growth were chosen based on the statistics for 'XHB', which is an ETF that tracks homebuilders within S&P500 [7]. The revenue/GM% correlation was estimated to be 5% as the bank expects there to be a small relationship between GM% and Revenue. The revenue/traded index correlation was estimated and a sensitivity analysis on the parameter follows. Table 7 displays the value of the flooring distributor using the MCF method Tool with varying estimates of the Revenue/traded index correlation. In the valuation, two of the MCF market parameters are adjusted to account for the liquidity premium. Both the risk-free rate and traded index expected growth were increased by 7%. As the MCF method does not incorporate liquidity risk, this is how an adjustment can be made to make the valuation consistent with the flooring distributor DCF.

Table 7. MCF method valuation of flooring distributor with varying correlation and terminal growth ($000s)

	Terminal Value Growth Rate				
Revenue/Traded Index Correlation	1.0%	1.5%	2.0%	2.5%	3.0%
0%	45,452	46,092	46,758	47,450	48,170
50%	44,792	45,416	46,064	46,738	47,439
100%	44,182	44,792	45,425	46,083	46,767

By comparing Table 6 and Table 7, it is clear that the two methods lead to similar valuations for the flooring distributor. This gives merit to the MCF method approach beyond its typical scope as solely an ROA tool. Not only are the valuations similar, but the sensitivity of the MCF method to its parameter estimates are markedly less. Additionally, the MCF method brings together the high, medium and low estimates for Revenue and GM% into a single valuation. The DCF fails to do so as it instead presents the results as three separate valuations. It is important to note that the MCF method also considers the spread between the high, medium and low cash flow estimates. That is, if the medium estimates are held constant, the high estimates raised, and the low estimates lowered proportionally, the MCF method will yield a lower valuation than the base case. This occurs because the resulting cash-flows are more uncertain and therefore, are essentially discounted at a greater rate. Another aspect of the MCF that was not featured here is that the growth rates of the low, medium and high revenues, which were assumed to be the same in all of the examples in Table 9, can be adjusted. For example, if the low, medium and high rates are chosen as -1%, 2% and 4% respectively, the base case scenario yields a valuation of $45,141. In this way, analysts can run many realistic scenarios to better hone in on the valuation.

4 Conclusions

Based on the results from Cases I and II, it is fair to assess the MCF method as a credible and useful valuation tool. The MCF method offers several advantages over standard DCF analysis:

- the parameters required by the MCF method are often less subjective than parameters found in DCF analysis,
- the application of real options to investments and the probability of success of the project is applied in a more natural way than a risk adjustment used in DCF analysis,
- the MCF method incorporates high, medium and low cash flow estimates into a single value while the DCF analysis is restricted to one output per scenario,
- the MCF method incorporates the divergence of the high and low cash flow estimates from the medium ones; it properly compares riskier projects to less risky based on the uncertainty of managerial specified cash-flow estimates.

In addition to the benefits listed above, the MCF tool developed as part of this work allows the investment bank analysts to apply the theoretical ROA framework with a user-friendly interface to better value investments.

References

[1] Trigeoris, L.: *Real Options: Managerial Flexibility and Strategy in Resource Alloscation*. The MIT Press, 1996.

[2] Copeland, P., and Tufano,T.: A real-world way to manage real options. *Harvard Business Review* **82** (2004), 90-99.

[3] Block, S.: Are Real Options actually used in the real world? *Engineering Economist* **52** (2007), 255-267.

[4] Bennouna, K., Meredith, G and Marchant, T.: Improved capital budgeting decision making: evidence from Canada. *Management Decision* **48** (2010), 225-247.

[5] Jaimungal, S., and Lawryshyn, Y.:Incorporating managerial information into real option valuation. Available at *SSRM: http://ssrn.com/abstract=1730355,* 2011.

[6] Black, F. and Scholes, M. The pricing of Options and Corporate Liabilities. *Journal of Political Economy*, (1973).

[7] ETF Database (2014). [Online]. Available: http://etfdb.com/. [Accessed 15 January 2014].

[8] Ochoa,C. M.: Valuation: from discounted cash flow (DCF) Approach to the Real Options Approach (ROA). *AD-MINISTER* **5** (2004), 59-73.

[9] Peskett,R.: Beyound DCF. *Management Accounting* **77**(1999), 60-61.

[10] Cassimon, D. and Engelen,P.:A compound option valuation model of R&D intensive pharmaceutical companies. *KORUS* **3** (2004), 209-212.

[11] Dixit, A. and Pindyck, R.: *Investment Under Uncertainty*. Princeton University Press, 1994.

[12] Newton, D., Paxson, D. and Widdicks, M.: Real R&D options. *International Journal of Management Reviews* **5** (2004), 113-130.

[13] Copeland, T. and Antikarov,V.: *Real options: A practitioner's guide*. Texere, 2001.

Numerical analysis of 3-D nonlinear financial model with lagged variables – influence of initial histories

Ladislav Lukáš [1]

Abstract. The paper concerns with numerical analysis of generalization of 3-D nonlinear financial model, which have been reported in the literature recently. It is composed of three simultaneous nonlinear ordinary differential equations with three state variables, which express the interest rate, the investment demand, and the price index, in particular. We have already introduced time dependent constitutive coefficient representing rate of investment instead of time invariant one. Further, we also admit more complex construction of right hand side terms including time delayed response of state variables, which yields the system with more general structure. In such case, the initial value problem turns to initial history problem, in general. We start with the simplest analytic form of the lagged terms, which assumes three intensities and three delays given. Numerical investigation of model with different forms of investment rate function and lagged structure is reported in detail. In particular, we are focused on analysis of influence of different initial history functions upon generated trajectories and their convergence to trajectories generated by corresponding initial value problems with neglected delays. All computations for solving initial value problems and initial history ones are performed by sw Mathematica.

Keywords: nonlinear financial model, initial value problem, lagged state variables, numerical solution.

JEL Classification: C63, G19
AMS Classification: 37N40, 91G80

1 Introduction

The present paper brings results of some numerical experiments done with the financial model we have already discussed earlier, see Lukáš [4]. There is a 3-D financial model which prototype was published in Ma [5], and Ma [6]. In general, the model is constituted from three state variables representing interest rate, investment demand, and price exponent, respectively, which are coupled together with their rate quantities by phenomenological evolution equations. Because of nonlinear coupling between investment demand rate and interest rate assumed, the model belongs to family of nonlinear equations which can produce attractive behaviors like bifurcations, chaotic behavior and strange attractors, too. We refer to Cai [1], Chen [2], and Yang [9] for further reading thereon.

Another effect which may produce more complicated behavior as well, there is a delay feedback of the system. In general case, a system of ordinary differential equations (ODE) turns into a system of delayed differential equations (DDE). In particular, we refer to Son [8] and Zhang [10] for some general aspects of formulation and handling delays in dynamic financial systems. An introduction to DDE theory may be found in Roussel [7], and numerical solution of DDE using Mathematica in [3] particularly

2 Model definition

Following Ma [5] and Lukáš [4], the model is given by equations (1) of DDE type which contain both nonlinear coupling and delay feedback.

$$\begin{aligned} \mathrm{d}x/\mathrm{d}t &= z + (y - a)x + k_1(x - x(t - \tau_1)),\ t \geq 0, \quad & x(t) &= u_0(t),\ t \in [-\tau_1, 0] \\ \mathrm{d}y/\mathrm{d}t &= 1 - by - x^2 + k_2(y - y(t - \tau_2)),\ t \geq 0, \quad & y(t) &= v_0(t),\ t \in [-\tau_2, 0] \\ \mathrm{d}z/\mathrm{d}t &= -x - cz + k_3(z - z(t - \tau_3))\ t \geq 0, \quad & z(t) &= w_0(t),\ t \in [-\tau_3, 0]. \end{aligned} \tag{1}$$

In total, it contains three non-negative constitutive parameters a, b, and c, which gauge saving amount, investment cost, and elasticity of demands of commercials transformed, respectively, further, three feedback intensities $k_j \in \mathrm{R}$, with delay times $\tau_j > 0$, j=1, 2, 3, called lags as well, given, and three initial histories $u_0(t)$, $v_0(t)$, $w_0(t)$,

[1] Univ. of West Bohemia, Faculty of Econ., Dpt. of Econ. and Quantitative Methods, Husova 11, 30614 Pilsen, lukasl@kem.zcu.cz.

given as well. In case of no feedback, i.e. for all $\tau_j = 0$, the model (1) turns to ODE one having initial values x_0, y_0, z_0 given by $x_0 = u_0(0)$, $y_0 = v_0(0)$, $z_0 = w_0(0)$, respectively.

All calculations were performed by Mathematica notebooks we developed for numerical investigation of model (1) using extremely powerful command `NDSolve`. Since the model (1) with its parameters and given data yields a plethora of possible combinations, we select just two topics and their combination to present now – *i*) numerical investigation of model with varied initial history functions, *ii*) analysis of influence of different delays on state-space trajectories.

3 Numerical experiments

First, we introduce another denotation of initial history functions $u_0(t)$, $v_0(t)$, $w_0(t)$, while we need to differentiate between original one and varied one, as well. Let denote $\boldsymbol{\eta}_o(t)$ a vector function representing an initial history functions, occasionally called as original history, too, and $\boldsymbol{\eta}_p(t)$ a vector function representing varied initial histories, called perturbed or varied history, as well. Solution of problem (1) with initial history $\boldsymbol{\eta}_o(t)$, or $\boldsymbol{\eta}_p(t)$, given will be denoted $\mathbf{x}_o(t)$, or $\mathbf{x}_p(t)$, respectively. Further, we restrict ourselves with the case of same delays in all three variables, thus simplifying the general form of feedback effects, now, introducing $(\tau_1, \tau_2, \tau_3)^T = (\tau_0, \tau_0, \tau_0)^T$.

3.1 Simple variation of constant initial history function

First example – the original history is given in form, which is conformal with initial data discussed already in Lukáš [4], and the varied history is chosen in form with just one initial history function changed, in particular, that one which represents increase in 30 percent of original level of investment demand applied. Both initial histories are given by (2). We apply also a=3, b=0.1, c=1, and k_1=0.1, k_2=0.2, k_3=0.1, respectively.

$$\boldsymbol{\eta}_o(t)=(\eta_{ox}(t), \eta_{oy}(t), \eta_{oz}(t))^T = (x0,y0,z0)^T =(2,3,2)^T ,$$

$$\boldsymbol{\eta}_p(t)=(\eta_{px}(t), \eta_{py}(t), \eta_{pz}(t))^T = (x0(1+x_p),y0(1+y_p),z0(1+z_p))^T =(2,3.9,2)^T , \boldsymbol{\Delta}_1 = (x_p, y_p, z_p)^T = (0, 0.3, 0)^T, \quad (2)$$

where $\boldsymbol{\Delta}_1$ is vector of variation constants given.

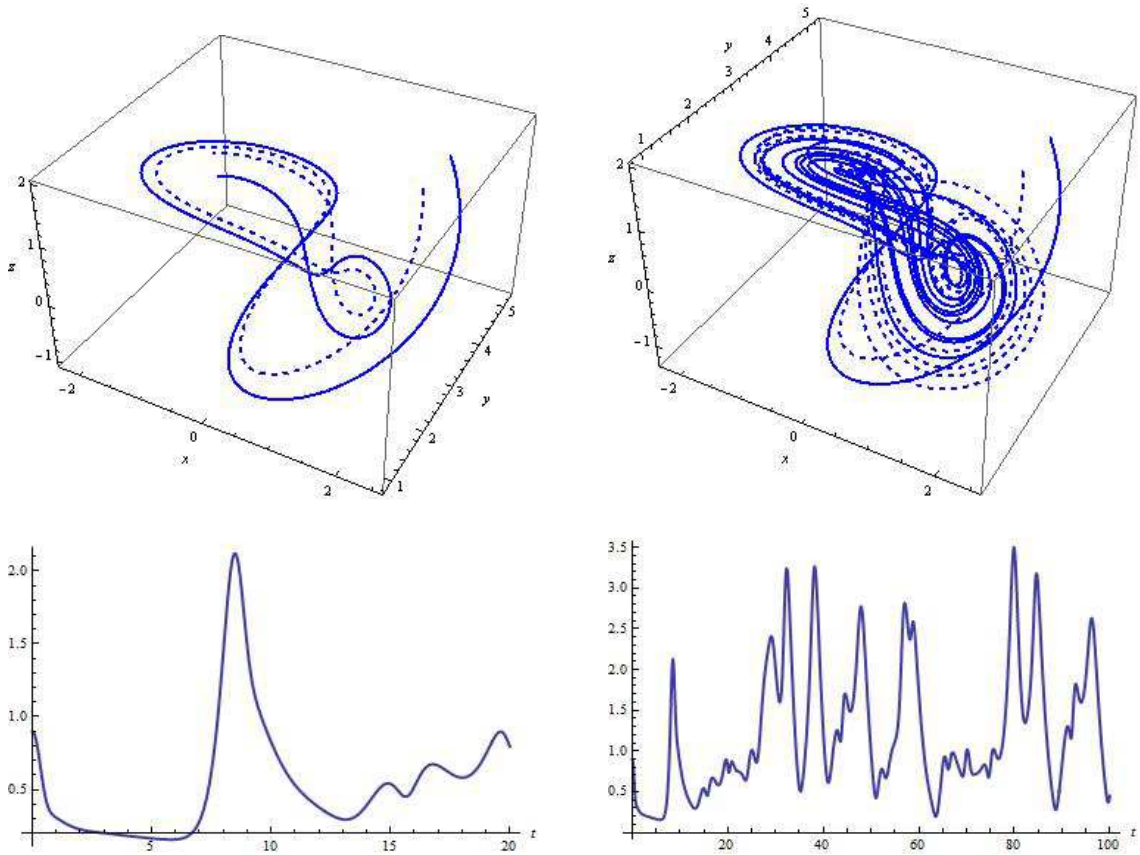


Figure 1 model (1)&(2), τ_0=0, solution $\mathbf{x}_o(t)$, $\mathbf{x}_p(t)$, distance $\delta(t)$; on left t_{max} = 20, on right t_{max} = 100

Fig. 1 shows computed trajectories for original history (dashed line) and varied history (thick line) with different t_{max}, t_{max} = 20 on the left, and 100 on the right. Bellow, there are functions $\delta(t)$ representing distance be-

tween solution with original history applied, denoted $\mathbf{x}_o(t)=(x_o(t), y_o(t), z_o(t))^T$, and solution with varied history applied, too, denoted $\mathbf{x}_p(t)=(x_p(t), y_p(t), z_p(t))^T$, which is given by (3)

$$\delta(t) = \sqrt{[(x_p(t) - x_o(t))^2 + (y_p(t) - y_o(t))^2 + (z_p(t) - z_o(t))^2]}. \quad (3)$$

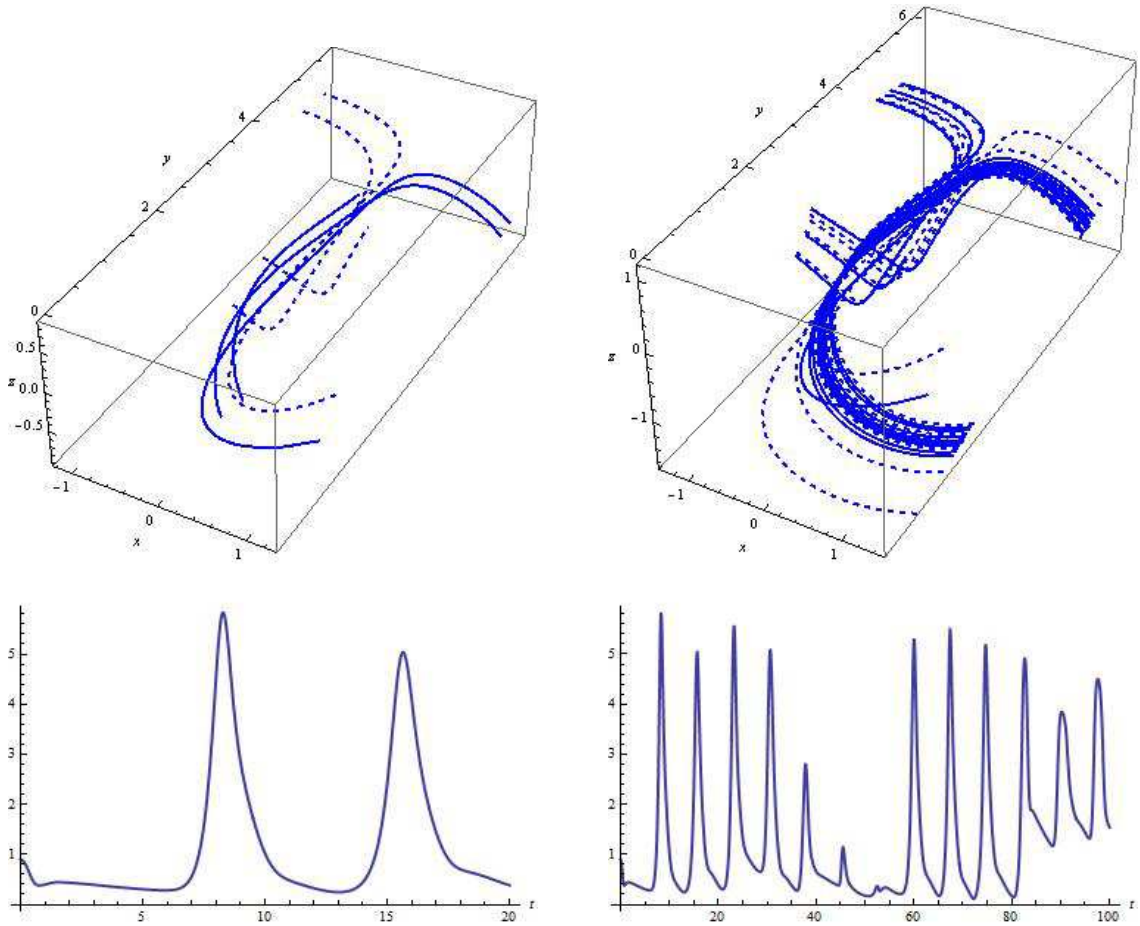
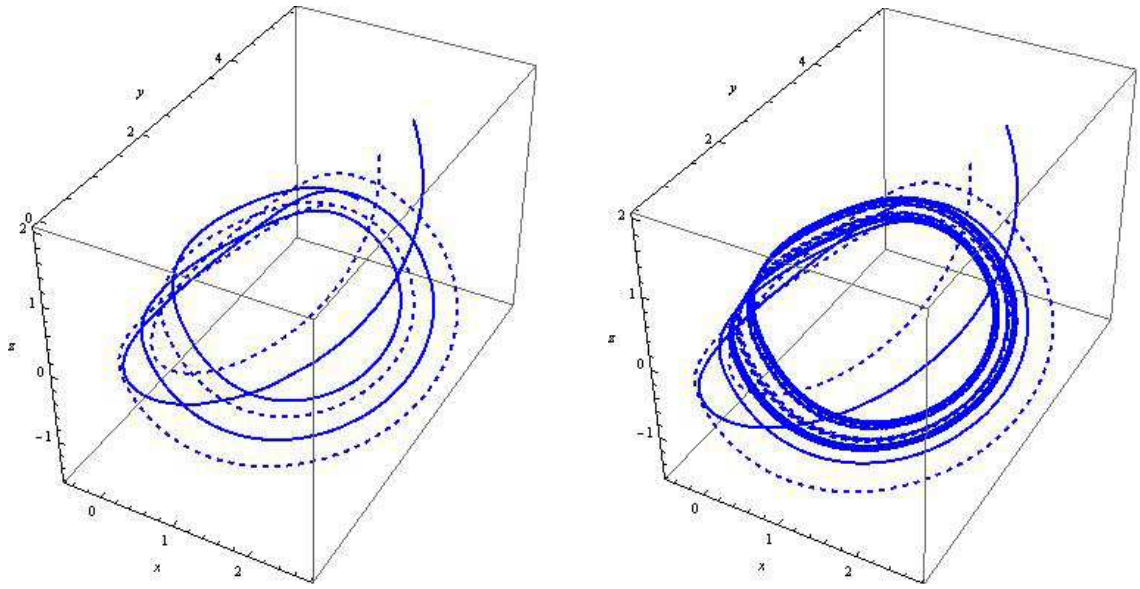


Figure 2 model (1)&(2), τ_0=1, solution $\mathbf{x}_o(t)$, $\mathbf{x}_p(t)$, distance $\delta(t)$; on left t_{max} = 20, on right t_{max} = 100

Fig. 2 shows computed $\mathbf{x}_o(t)$, $\mathbf{x}_p(t)$, and distance $\delta(t)$, but with τ_0=1. We can see that trajectories have changed dramatically as being compared with those ones depicted on Fig. 1, where it was τ_0=0. Comparing distances, in particular the cases with t_{max} = 100, we may conclude that on Fig. 1 it looks more chaotic than on Fig. 2, where it resembles a type of 'quasi-periodic' behavior, at least for the solution time span t_{max} = 100.



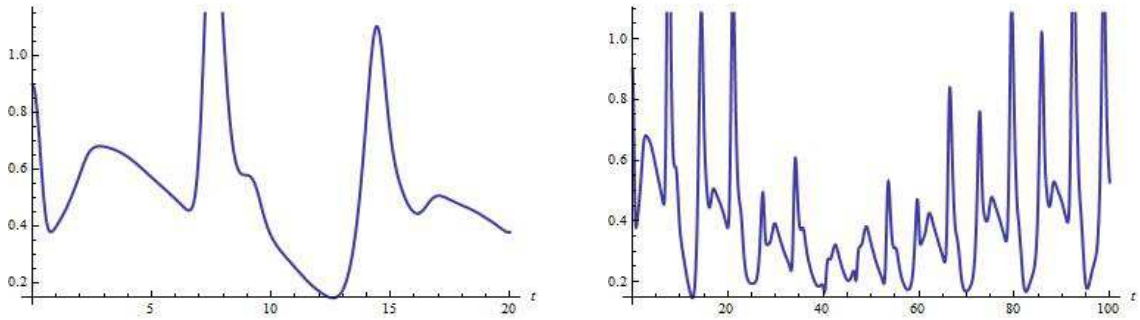


Figure 3 model (1)&(2), τ_0=2, solution $\mathbf{x}_o(t)$, $\mathbf{x}_p(t)$, distance $\delta(t)$; on left t_{max} = 20, on right t_{max} = 100

Fig. 3 brings similar results. However, we may conclude that trajectories have changed again when taking τ_0=2, though not so much as for τ_0=1. On the opposite, $\delta(t)$ looks to be more disturbed than in the previous case.

3.2 Damped cyclic variation of initial history functions

Second example – the histories $\boldsymbol{\eta}_o(t)$ and $\boldsymbol{\eta}_p(t)$ are chosen in much sophisticated way. We are using the same analytic form for both ones but with different constants given – see (4a,b). They share the same initial values at t=0, i.e. $\boldsymbol{\eta}_o(0) = \boldsymbol{\eta}_p(0) = (4,6,4)^T$, with $(x0,y0,z0)^T = (2,3,2)^T$ being given. However, they differ on initial history time period $[-\tau_h,0]$, where $\tau_h = \tau_0 = 2$. We use a=3, b=0.1, c=1, and k_1=0.1, k_2=0.2, k_3=0.1, again.

$$\boldsymbol{\eta}_o(t)=(\eta_{ox}(t), \eta_{oy}(t), \eta_{oz}(t))^T = (x0(1+\exp(e_{ox}t)\cos(c_{ox}t)),y0(1+\exp(e_{oy}t)(1+\sin(c_{oy}t)),z0(1+\exp(e_{oz}t)\cos(c_{oz}t)))^T,$$

$$(e_{ox}, e_{oy}, e_{oz})^T = (0.881797, 0.222095, 1.0327)^T\ ,\ (c_{ox}, c_{oy}, c_{oz})^T = (9.80871, 3.95561, 4.17592)^T\ , \qquad (4a)$$

$$\boldsymbol{\eta}_p(t)=(\eta_{px}(t), \eta_{py}(t), \eta_{pz}(t))^T = (x0(1+\exp(e_{px}t)\cos(c_{px}t)),y0(1+\exp(e_{py}t)(1+\sin(c_{py}t)),z0(1+\exp(e_{pz}t)\cos(c_{pz}t)))^T,$$

$$(e_{px}, e_{py}, e_{pz})^T = (e_{ox}+\varepsilon_x, e_{oy}+\varepsilon_y, e_{oz}+\varepsilon_z)^T = (1.23904, 0.249672, 0.962755)^T\ , \qquad (4b)$$

$$(c_{px}, c_{py}, c_{pz})^T = (c_{ox}+\gamma_x, c_{oy}+\gamma_y, c_{oz}+\gamma_z)^T = (9.99387, 3.87506, 3.79381)^T,$$

$$\boldsymbol{\Delta}_{2e}=(\varepsilon_x, \varepsilon_y, \varepsilon_z)^T=(0.357241, 0.02757653, -0.069942)^T,\ \boldsymbol{\Delta}_{2c}=(\gamma_x, \gamma_y, \gamma_z)^T=(0.18516, -0.0805487, -0.382116)^T.$$

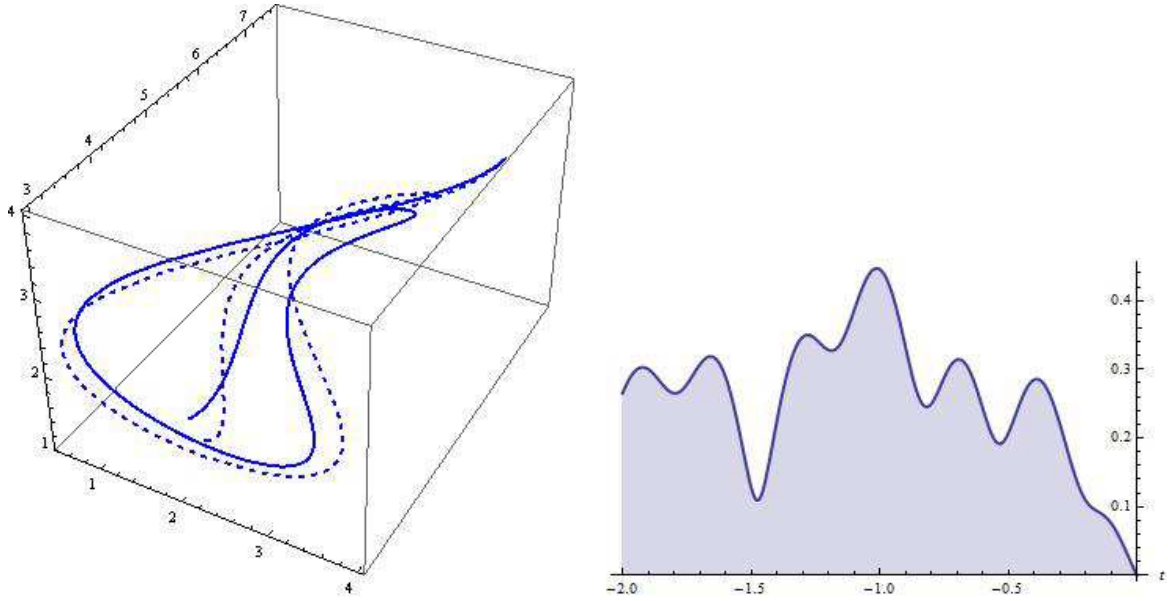


Figure 4 initial histories $\boldsymbol{\eta}_o(t)$ (dashed line), $\boldsymbol{\eta}_p(t)$ (thick one) and its distance $\delta(t)$ on $[-\tau_h,0]$, τ_h = 2

Fig. 4 shows the initial histories defined by (4a,b) containing a lot of constants. Since $\boldsymbol{\eta}_o(0) = \boldsymbol{\eta}_p(0)$ being selected, the solutions $\mathbf{x}_o(t)$, $\mathbf{x}_p(t)$ will not be different ones unless τ_0>0. Hence for illustration, we choose τ_0=0.5, first, and at second, τ_0=1.5, thus investigating cases with fractional time lags, too.

Fig. 5 gives the results with τ_0=0.5. Interesting is that for t_{max} = 20 the trajectories $\mathbf{x}_o(t)$, $\mathbf{x}_p(t)$ seem to be almost the same. However, they are different ones as it becomes evident by inspection their distance function $\delta(t)$, which shows the maximum value about 0.08 for all t ranging from 0 to 20, even till 30, as it is recognizable when looking at $\delta(t)$ for t within period [0,100].

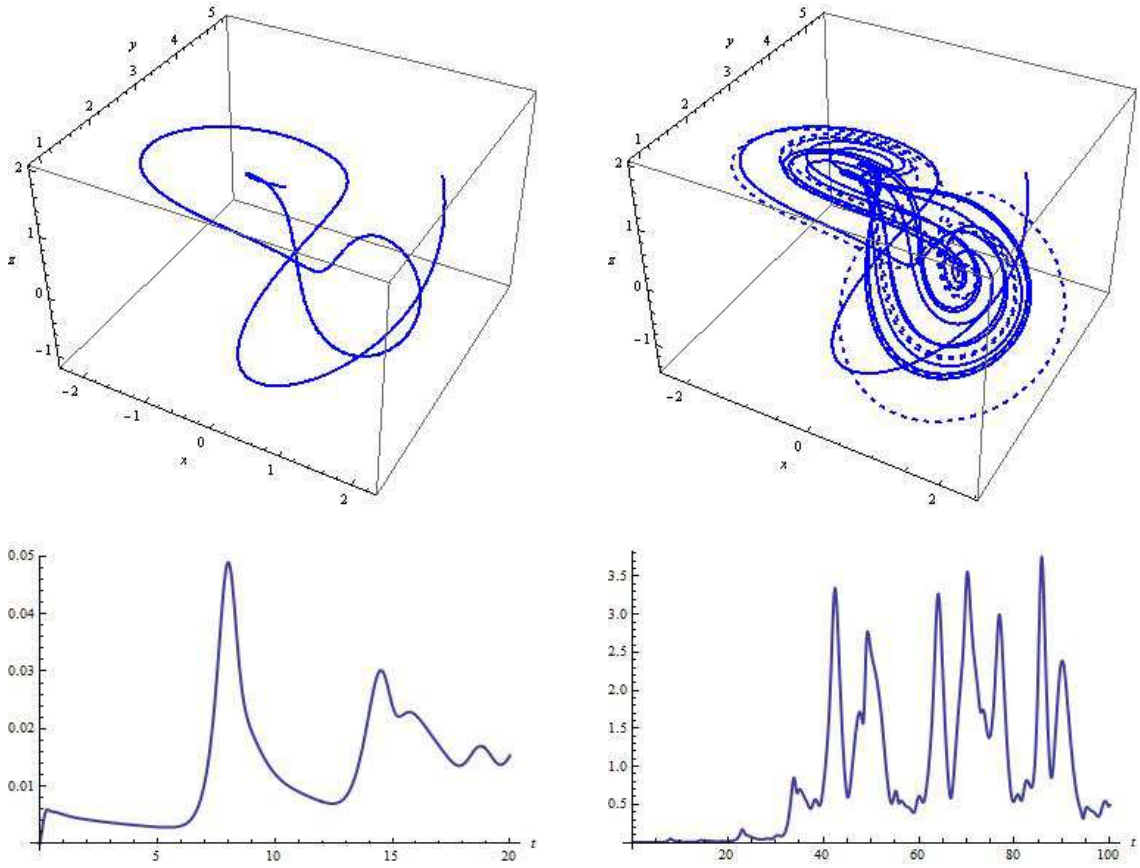


Figure 5 model (1)&(4a,b), τ_0=0.5, solution $\mathbf{x}_o(t)$, $\mathbf{x}_p(t)$, distance $\delta(t)$; on left t_{max} = 20, on right t_{max} = 100

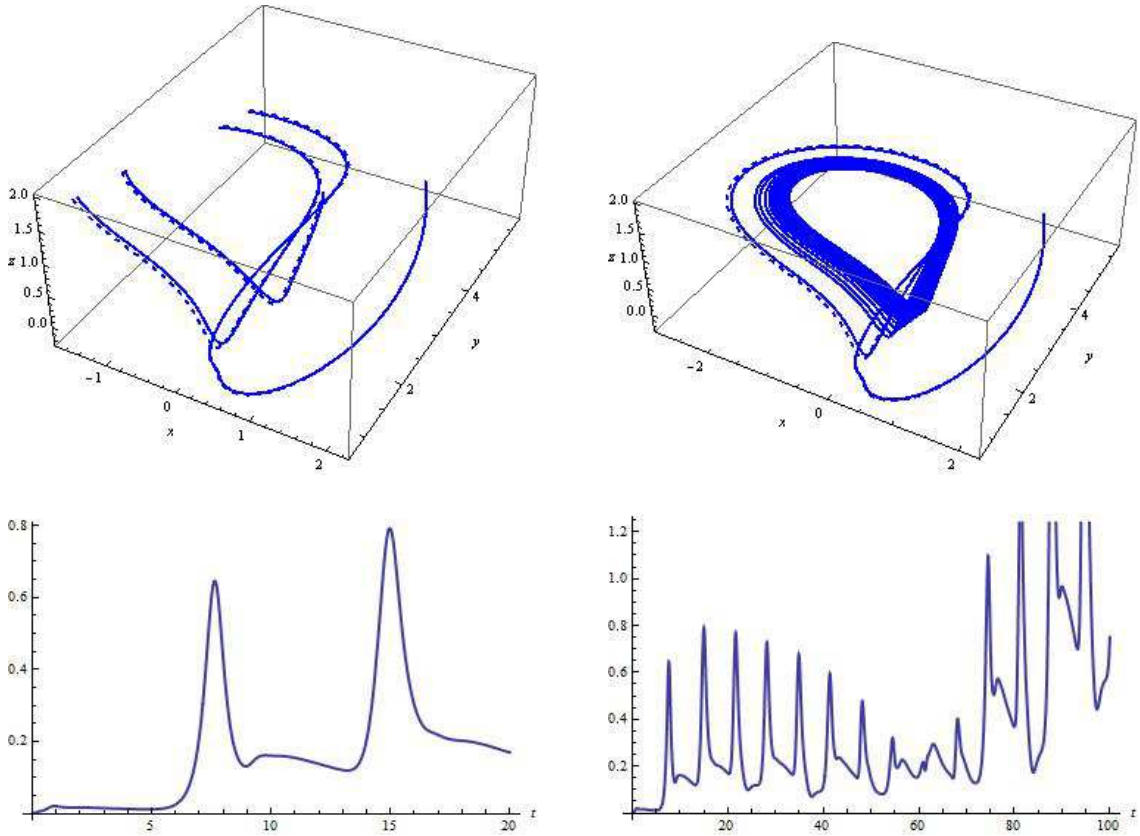


Figure 6 model (1)&(4a,b), τ_0=1.5, solution $\mathbf{x}_o(t)$, $\mathbf{x}_p(t)$, distance $\delta(t)$; on left t_{max} = 20, on right t_{max} = 100

Though different, the trajectories computed with different models and depicted on Fig. 1 and Fig. 5 show some common evolution features, which is remarkable for $t_{max} = 100$, in particular. In this case, one can conclude that influence of short lag τ_0=0.5 and relatively simple history on $[-\tau_0,0]$ given looks to be rather weak.

Now, taking τ_0=1.5 (three times longer lag and the initial history is more complicated than on Fig. 5 before), we get trajectories which are depicted on Fig. 6. They are different ones already just by inspection. After a relatively short time elapsed till $t < 6$, as being evident on both figures depicting trajectories for $t_{max} = 20$, and $t_{max} = 100$, and on their distances $\delta(t)$, in particular, both solutions show some kind of loop oscillations interrupted at vicinity of t =60 suddenly, and going further in similar loops in 3-D state-space, but yielding highly disturbed distance picture. It becomes evident through function values of $\delta(t)$ for t ranging from 70 till $t_{max} = 100$, and prospectively even larger one.

Finally, inspecting trajectories depicted on Fig-s. 1, 2, 3, 5 and 6, we are able to see various types of complicated cycles both for perturbed and unperturbed initial histories, though some orbits are incomplete as being exported by Mathematica `ParametricPlot3D` command with default output options used.

4 Conclusion

We have discussed just two interesting features which occur in the selected nonlinear 3-D financial model with lagged variables analyzed numerically within a framework of DDE theory.

First, there is an investigation of influence of different initial history functions on distance between solutions, or say in another words, on stability of solution. The field of prospective investigations therein is pretty large.

The second feature analyzed was focused on investigation of influence of feedback delays, and we presented just two case studies thereon. Particularly interesting, there is coupling of both effects.

Hence, our further research in this field will be focused particularly on

- complex program of numerical investigation of large parametric space of the presented model;
- advanced theoretical development including analysis of bifurcations;
- empirical verification of the model and estimation of their constitutive parameters.

Acknowledgements

The research project was supported by Grant SGS-2014-047 of the Faculty of Economics, University of West Bohemia in Pilsen.

References

[1] Cai, G., Huang, J.: A New Finance Chaotic Attractor. *Int.J. Nonlinear Science*, **3** (2007), 3, 213-220. http://worldacademicunion.com/journal/1749-3889-3897IJNS/IJNSVol3No3Paper08 .

[2] Chen, W.-C.: Nonlinear dynamics and chaos in a fractional-order financial system. *Chaos Solitons Fractals*, **36** (2008), 1305-1314.

[3] Delay Differential Equations. Mathematica Tutorial. https://reference.wolfram.com/mathematica/tutorial/NDSolveDelayDifferentialEquations.html.

[4] Lukáš, L.: Numerical investigation of 3-D nonlinear financial model. In *Proceedings of* 31-*st Int.Conf. Mathematical Methods in Economics* 2013. (Vojáčková, H., editor.). College of Polytechnics Jihlava, Jihlava, 2013, 546-551.

[5] Ma, J.-H., Chen, Y.-S.: Study for the bifurcation topological structure and the global complicated character of a kind of nonlinear finance system (I). *Appl. Math. Mech.-Engl.* **22** (2001), 11, 1240-1251.

[6] Ma, J.-H., Chen, Y.-S.: Study for the bifurcation topological structure and the global complicated character of a kind of nonlinear finance system (II). *Appl. Math. Mech.-Engl.* **22** (2001), 12, 1375-1382.

[7] Roussel, M.R.: Delay-differential equations. http://people.uleth.ca/~roussel/nld/delay.pdf.

[8] Son, W.-S., Park, Y.-J.: Delay effect on the dynamical model of financial system. http://www.phys.sinica.edu.tw/~socioecono/econophysics2010/pdfs/SonPaper.pdf .

[9] Yang, M., Cai G.: Chaos Control of a Non-linear Finance System. *JournalUncertain Systems*, **5** (2011), 4, 263-270. http://www.worldacademicunion.com/journal/jus/jusVol05No4paper03.pdf .

[10] Zhang, R.: Bifurcation Analysis for a Kind of Nonlinear Finance System with Delayed Feedback and Its Application to Control of Chaos. Hindawi Publ.Corp., *J.Appl.Math.*, 2012, Article ID 316390, 18 pp.

Multivariate approach to testing VaR models

Marta Malecka[1]

Abstract. Evaluation of *VaR* models has traditionally focused on *VaR* failure process at one tolerance level, therefore it has been restrained to verifying the accuracy of predictions of one chosen quantile of the return distribution. The additional information can be used by testing the goodness of fit of the entire return distribution. Such approach implies using equal weights for the interior and the tails of the distribution, which may not correspond with risk management objectives. A compromising approach is to apply a multivariate *VaR* test aimed at testing the *VaR* model at multiple *VaR* levels.

In the paper we proposed a new test for multiple *VaR* levels, based on multivariate correlation matrix. Through the simulation study we compared its performance to the multivariate Ljung Box *VaR* test. The results of the study showed that the proposed approach outperformed the Ljung Box test in terms of both size and power for the series length of 250 observations.

Keywords: Multivariate *VaR*, correlation matrix test.

JEL classification: C22, C52, D53, G11
AMS classification: 62M10, 91B84, 62P05

1 Introduction

The idea behind *VaR* model evaluation is to use the observed failure process to assess whether the failure rate matches the assumed level of tolerance and whether the failures are not serially correlated. Verification of these two criteria by means of the univariate tests is limited to the analysis of the failure process for one, chosen tolerance level. The univariate failure process is defined as stochastic process of the indicator variable, which, at each time point, informs whether the VaR level was exceeded. Such analysis checks the accuracy of *VaR* predictions through one quantile of rate of return distributions, equal to the chosen tolerance level.

The information about the whole return distribution can be used in *VaR* model evaluation through testing the goodness of fit of the predicted return distribution to the observed return trajectory. Such a general approach implies using equal weights for the interior and the tails of the distribution, which may not be relevant to the management of extreme risk.

A compromising approach is based on a stochastic process of a multivariate indicator variable, which involves multiple levels of tolerance. Through the multivariate approach statistical tests simultaneously verify *VaR* models at chosen number of tolerance levels. Thus the model evaluation covers the custom area of the return distribution and usually concentrates on the tails of the distribution. To jointly test the failure process for different tolerance levels Hurlin Tokpavi [4] proposed the use of the test statistic of Li & McLeod [5], which is a multivariate extension to the Ljung-Box test [3]. Following this line of research, the paper focuses on the multivariate tests aimed at detecting serial correlation in *VaR* failures. We adopted the approach based on correlation matrices and compared it to the multivariate *VaR* test based on the Ljung-Box statistic.

The aim of the paper was to present the proposition of a test for multiple *VaR* levels, based on multivariate correlation matrix test framework. Through the simulation study, the statistical properties of the test were compared to the multivariate Ljung-Box test for multiple *VaR* levels. The size and power assessment was based on the GARCH process assumption.

[1]University of Lodz, Department of Statistical Methods, 41 Rewolucji 1905 r., Lodz, Poland, marta.malecka@uni.lodz.pl

2 VaR multivariate correlation test

VaR belongs to the group of quantile risk measures, hence its definition is based on the notion of the upper p-quantile. Let us consider the probabilistic space $(\Omega, \mathcal{F}, P)$ and the random variable $X : \Omega \longrightarrow \mathbb{R}$. Value-at-risk ($VaR$) at the level of tolerance $p \in (0,1)$ of the variable X is defined as

$$VaR_p(X) = -q^p(X), \tag{1}$$

where $q^p(X)$ is the upper p-quantile of X,

$$q^p(X) = \inf\{x \in \mathbb{R} : P(X \leq x) > p\}.$$

For the continuous random variables with strictly increasing distribution function, the definition of VaR reduces to the p-quantile of the X distribution

$$VaR_p(X) = F_X^{-1}(p). \tag{2}$$

The verification of the hypothesis regarding serial correlation of VaR failures can be done through the multivariate test for correlation matrices, which is designed to examine the independence of random subvectors. The procedure is based on the assumption that the vector of all observed variables can be divided into subvectors of related variables. It is then checked whether the variables that belong to the given subvector are independent of all other variables in all other subvectors. The test construction is based on the likelihood ratio statistic and uses estimators of correlation matrix, divided into blocks that refer to different subvectors.

Let $\mathbf{X}$ be the normally distributed random vector of n random variables with the mean μ and covariance matrix $\Sigma : \mathbf{X} \sim \mathbf{N}_n(\mu, \Sigma)$. Let us assume that the vector $\mathbf{X}$ can be divided into $H+1$ subvectors of related variables:

$$\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1, ..., \mathbf{X}_H). \tag{3}$$

The covariance matrix Σ can be written in the following block form

$$\Sigma = \begin{bmatrix} \Sigma_{00} & \Sigma_{01} & ... & \Sigma_{0H} \\ \Sigma_{10} & \Sigma_{11} & ... & \Sigma_{1H} \\ \vdots & \vdots & & \vdots \\ \Sigma_{H0} & \Sigma_{H1} & ... & \Sigma_{HH} \end{bmatrix}. \tag{4}$$

Let us denote m_h the number of variables in the subvector X_h, $h = 0, 1, ..., H$, $m_0 + m_1 + ... + m_H = n$. Under the above notation $(\mathbf{X}_0, \mathbf{X}_1, ..., \mathbf{X}_H)$ does not represent a random sample, but it represents the composition of the random vector. Let us consider the hypothesis

$$H_0 : \Sigma_{gh} = \mathbf{0}, \quad g, h \in \{0, ..., H\}, \quad g \neq h, \tag{5}$$

stating the zero covariance of the vectors belonging to different groups, which can be alternatively written as

$$H_0 : \Sigma = \begin{bmatrix} \Sigma_{00} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \Sigma_{11} & ... & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & ... & \Sigma_{HH} \end{bmatrix}. \tag{6}$$

Adopting the multivariate correlation test framework to VaR model evaluation requires the assumption that the vector of all observed variables includes series of VaR failures at all chosen tolerance levels in all considered moments. This vector needs to be divided into subvectors grouping VaR failures at different tolerance levels, observed at the same moment.

Let $I_t(p) = \mathbf{1}_{\{-X_t > VaR_p(X_t)\}}$ denote the VaR failure process for the variables $X_t : \Omega_t \longrightarrow \mathbb{R}$ at the level of tolerance $p \in (0,1)$, $t = 1, ..., T$, and let us assume that $\mathbf{p} = (p_1, p_2, ..., p_m)$ is the series of m considered tolerance levels. Let us define the Hit process

$$Hit_t(p) = I_t(p) - p, \quad t = 1, ..., T. \tag{7}$$

In *VaR* model evaluation, the vector of observed variables takes the form

$$Hit = (Hit_T(p_1), Hit_T(p_2), ..., Hit_T(p_{m_0}), ..., Hit_{T-H}(p_1), Hit_{T-H}(p_2), ..., Hit_{T-H}(p_{m_H})). \quad (8)$$

The variables need to be grouped in subvectors

$$Hit_h = (Hit_{T-h}(p_1), Hit_{T-h}(p_2), ..., Hit_{T-h}(p_{m_h})), \quad (9)$$

which include one moment $T-h$ and all chosen tolerance levels, where $h \in \{0, 1, ..., H\}$, $1 \leq H < T$ is a fixed lag order.

VaR model evaluation criteria imply [1]

$$H_0 : cov(Hit_T Hit_{T-h}) = 0 \quad \forall h = 1, 2, ..., H, \quad (10)$$

which can be written through the covariance matrix of vector Hit as

$$H_0 : cov(HitHit) = \begin{bmatrix} cov(Hit_T Hit_T) & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & cov(Hit_{T-1} Hit_{T-1}) & ... & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & ... & cov(Hit_{T-H} Hit_{T-H}) \end{bmatrix}. \quad (11)$$

The above formula is equivalent to (6), where $\Sigma_{hh} = cov(Hit_{T-h} Hit_{T-h})$, $h = 0, 1, ..., H$.

The likelihood ratio statistic of the above hypothesis has the form

$$LR_H = \frac{\det \tilde{\mathbf{R}}}{\det \hat{\mathbf{R}}_{00} \det \hat{\mathbf{R}}_{11} \cdots \det \hat{\mathbf{R}}_{HH}}, \quad (12)$$

where $\hat{\mathbf{R}}_{hh}$, $h = 0, 1, ..., H$ is the sample correlation matrix for the subvector Hit_{t-h} and $\tilde{\mathbf{R}}$ is the estimate of the correlation matrix of the vector Hit. Hence $\mathbf{R}_{gh} = \mathbf{0}$, $g, h = 0, 1, ..., H$, $g \neq h$. It can be shown that for $H = 1$ the above statistic has the Wilks distribution [7]. In case of non-normality of the observed random vector, analogously to the Ljung-Box test, the statistical inference is based on asymptotic theorems [5]. If $H > 1$ the modified statistic can be used

$$LR_H^* = (n-1)c \ln LR_H, \quad (13)$$

where

$$c = 1 - \frac{1}{12b(n-1)}(2a_3 + 3a_2),$$

$$b = \frac{a_2}{2}, \quad a_2 = n^2 - \sum_{h=0}^{H} m_h^2, \quad a_3 = n^3 - \sum_{h-1}^{H} m_h^3,$$

n is the sample size.

The number of parameters in the matrix Σ is $\frac{1}{2}n(n+1)$ and under H_0 the number of parameters in all matrices Σ_{hh} equals to $\frac{1}{2}m_h(m_h+1)$. This gives the total number of the parameters in the restricted model equal to $\frac{1}{2}\sum_{h=1}^{H} m_h(m_h+1)$. Thus the number of degrees of freedom, equal to the difference between the number of parameters in the unrestricted and the restricted model, is

$$\frac{1}{2}n(n+1) - \frac{1}{2}\sum_{h=0}^{H} m_h(m_h+1) = \frac{a_2}{p} = b.$$

The distribution of LR_H^* can therefore be approximated by the χ_b^2 distribution [7].

In risk management, it can be presumed that the choice of the analyzed tolerance levels does not depend on the lag order of the *VaR* failure variables. If in all subvectors Hit_{t-h}, $h = 0, 1, ..., H$, $t \in \{1, ..., T\}$, the same tolerance levels are considered, the examined subvectors have equal length: $m_0 = m_1 = ... = m_H = m$. Then the relevant elements in formula (13) reduce to:

$$b = \frac{a_2}{2}, \quad a_2 = n^2 - (H+1)m^2, \quad a_3 = n^3 - (H+1)m^3.$$

3 Size and power evaluation

The size and power of the proposed LR_H test were compared to the properties of the multivariate Ljung-Box correlation test LB_H through the simulation study. The simulation experiment was based on the GARCH process, which guaranteed representation of the volatility clustering phenomenon. The size and power were estimated as the proportion of type-one errors and the proportion of the rejections under the alternative respectively. The estimates were computed over 10000 replications for sample sizes $T = 250, 500, 750, 1000$. The level of significance was set to 5%. The size evaluation was complemented by the graphical analysis, where the shapes of the empirical and theoretical distributions of the test statistics were compared.

The multivariate failure process in the simulation study was obtained in a two-step procedure. Firstly the return data was generated and then *VaR* estimates on two levels of tolerance were calculated, which gave the two-dimensional failure indicator variable at each time point. In the size exercise the return data $R_t,\ t = 1, ..., T$ was generated from the $GARCH(1,1)$ process given by:

$$\begin{aligned} R_t &= \sigma_t Z_t, \quad Z_t \sim N(0,1), \\ \sigma_t^2 &= \omega_1 + \alpha_1 R_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \tag{14}$$

where $\omega_1 = 0,000001$, $\alpha = 0,14$, $\beta_1 = 0,85$ [1]. *VaR* estimates were fixed at the levels of the 0.01– and 0.05–quantile of the R_t distributions, $t = 1, ..., T$. The two resulting series of *VaR* failures were constructed from independent observations and characterized by the ratio of *VaR* exceedances corresponding to the 1% and 5% tolerance levels.

The graphical analysis of the discrepancies between the empirical and theoretical distributions of the test statistics was done through the comparison of the shapes of the densities and cumulative distribution functions. $M = 10000$ values of the test statistics $S_1, S_2, ..., S_M$ were simulated and the empirical distribution function was computed as [2]

$$\hat{F}_M(x) = \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}_{S_i \leq x},$$

where $S_i \in \{LB_{Hi}, LR_{Hi}\}$ denote the $i-$th value of the evaluated test statistic LB_H or LR_H.

In case of both tests – the Ljung-Box LB_H test and the proposed LR_H test – the size estimates were over 10% for series of 250 observations, with the size estimate for LB_H being close to 20% (Table 1). The decreasing tendency in the probability of the type-one error gave size estimates below 10% for both tests for the series of 1000 observations. The discrepancies between the empirical and theoretical distribution shapes were confirmed by the graphical analysis (Figure 1– 2).

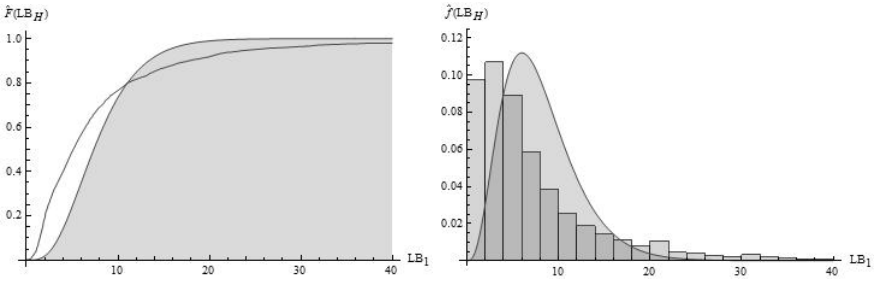
Test	Series length			
	250	500	750	1000
LB_H	0,199	0,126	0,105	0,092
LR_H	0,102	0,092	0,095	0,087

Table 1 Size estimates of multivariate *VaR* tests

The comparison of the estimated LB_H and LR_H test sizes indicated that, with both tests being above 5%, the proposed LR_H statistic gave results closer to the nominal significance level. Especially for the shortest considered series – of 250 observations – the LB_H size estimate was around 20%, while employment of the LR_H statistic reduced the probability of the type-one error to nearly 10%.

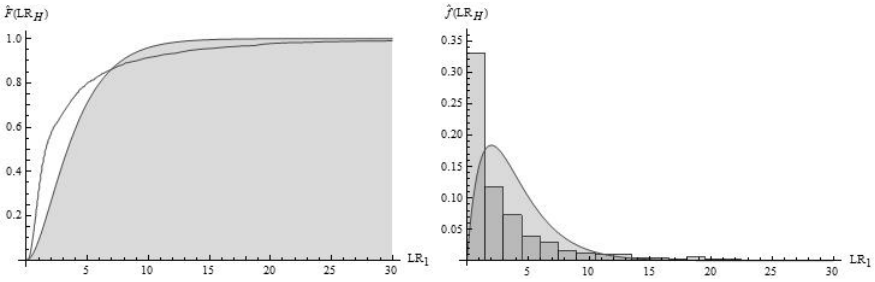
The power assessment required generating serially correlated failure process. The return data was generated from the GARCH process (14), which represents the volatility clustering phenomenon, while VaR estimates were set to the constant level. The experiment utilized the property of the GARCH process, where the volatility clustering can be measured and controlled by the correlation coefficient of the squared returns. The GARCH parameters ω_1 and β_1 were set to $\omega_1 = 0,000001$, $\beta_1 = 0,85$. The

[1] The parameter values were fixed on the basis of the initial study for six stock market indexes [6].



(a) cdf (b) pdf

Figure 1 Empirical and theoretical distribution of LB_H for 250 observations



(a) cdf (b) pdf

Figure 2 Empirical and theoretical distribution of LR_H for 250 observations

value of α_1 was chosen in a way that gave the correlation of squared returns ρ_1 subsequently on levels 0,1; 0,3; 0,5. *VaR* estimates were constant over time and fixed at the levels of the 0.01– and 0.05–quantile of the marginal distribution of the rate of return R. The algorithm produced two series of dependent *VaR* failures, whose overall shares in the observations corresponded to the assumed levels of tolerance.

In the power comparison we employed the Monte Carlo test technique, which is based on the assumption that the finite sample distribution of the test statistic can be simulated [2]. According to the Monte Carlo test algorithm we used the test statistic distributions simulated under H_0, which ensured the exact test size and guaranteed the comparability of the power estimates.

Test	ρ_1	Series length			
		250	500	750	1000
LB_H	0,1	0,05	0,32	0,53	0,94
	0,3	0,06	0,73	0,92	0,98
	0,5	0,07	0,85	0,97	0,99
LR_H	0,1	0,18	0,27	0,34	0,39
	0,3	0,35	0,58	0,70	0,81
	0,5	0,43	0,66	0,78	0,85

Table 2 Power estimates of multivariate *VaR* tests

The comparison of the power estimates between the statistic LB_H and the proposed statistic LR_H showed that the proposed approach allowed for the reduction of the probability of the type-two error in case of the shortest examined time series (Table 2). For the series of 250 observations the LB_H rejection frequencies were between 5% and 7%, depending on the strength of the correlation. The relevant frequencies for LR_H were over 18% and reached 43% in case of strongest correlation. For longer series the differences in power estimates between the tests were less evident. The results for different sample sizes showed that, in case of examined alternatives to H_0, the LB_H statistic exhibited faster convergence

to the theoretical distribution.

The superiority of the LR_H test in case of shortest time series may be regarded very important in business practice, as 250 observations is approximately equivalent to the one year of daily data. Such series length is also advocated by the Basel Committee on Banking Supervision in analysis related to market risk management.

4 Summary and conclusions

The study investigated correlation tests for *VaR* model evaluation. The focus was on the multivariate approach, which allows for comprehensive analysis of *VaR* model accuracy, covering simultaneously a chosen number of tolerance levels. The multivariate test based on correlation matrices was adopted to detect correlation in a multivariate VaR failure process.

Through the simulation study the size and the power of the proposed test were compared to the properties of the multivariate Ljung-Box test for serial correlation. The study showed the superiority of the proposed approach, in terms of both size and power, for the series length of 250 observations, which is the sample size recommended by the Basel Committee on Banking Supervision. For this series length the type-one error probability for the proposed test was reduced twice in comparison to the Ljung-Box procedure. Type-two error probabilities for the proposed test, in case of this series length, were several times lower than for the test based on the Ljung-Box statistic.

Acknowledgments

The research was supported by the Polish National Science Centre grant UMO-2013/11/N/HS4/03354.

References

[1] Campbell, S. D.: A Review of Backtesting and Backtesting Procedures, *Finance and Economics Discussion Series* Divisions of Research and Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C, 2005.

[2] Dufour, J. M.: Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics, *Journal of Econometrics* **133(2)** (2006), 443–477.

[3] Ljung, G. M., and Box G. E. P.: On a Measure of Lack of Fit in Time Series Models, *Biometrika* **65** (1978), 297–303.

[4] Hurlin, Ch., and Tokpavi, S.: Backtesting value-at-risk accuracy: a simple new test, *The Journal of Risk* **9(2)** (2007), 19–37.

[5] Li, W. K., and McLeod, A. I.: Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models, *Journal of the Royal Statistical Society serie B* **43** (1981), 231–239.

[6] Malecka, M.: Prognozowanie zmiennoci indeksw giedowych przy wykorzystaniu modelu klasy GARCH, *Ekonomista* **6** (2011), 843–860.

[7] Rencher, A. C.: *Methods of Multivariate Analysis*. John Wiley & Sons Inc., New York, 2002.

ICT and technical efficiency of 15 EU countries: A stochastic frontier analysis approach

Jan Manďák[1]

Abstract. The purpose of this paper is to assess the technical efficiency of 15 countries of the European Union (EU) and to investigate factors, which influence inefficiency using Stochastic Frontier Analysis approach during the time period 1995 - 2007. A Cobb-Douglas stochastic production frontier is simultaneously estimated with a technical inefficiency model. The impact of ICT (Information and Communication Technologies) on labour productivity and the effects of ICT in reducing aggregate technical inefficiency are analyzed. Countries of EU are also compared and ranked according to their levels of efficiency. Strong evidence is provided for significantly positive ICT impact on labour productivity. Further evidence indicates positive impact of ICT and technical education in reducing country inefficiencies. The most efficient countries in the sample are France, Italy and Netherlands followed by the other EU-12 countries. The obtained results are comparable with similar study Dimelis and Papaioannou [3], but it seems that in the EU trade openness has no impact in reducing country inefficiencies.

Keywords: stochastic frontier analysis, efficiency, ICT, EU, inefficiency model

JEL Classification: O47
AMS Classification: 91G70

1 Introduction

At least in last twenty years, Information and Communication Technologies (ICT) are considered as a main economic driver in many sectors of the economy. Undoubtedly, ICT had a significant impact on economic growth for both the United States and the European Union (Van Ark et al. [13]). According to the OECD [7] or Quiang [9], three main channels can be identified through which ICT can affect potential growth rates:

- an acceleration of productivity in the ICT-producing sector themselves, and a growing size of ICT-producing sectors in the economy;
- capital deepening across the economy, driven by rapid investment in ICT equipment, and resulting in a boost to labour productivity;
- widespread spillover effects on productivity arising from ICT.

Stiroh [11] states, that the majority of empirical studies have identified the ICT effect as an impact that comes from the traditional channel of capital deepening. However, it is also important to examine whether a part of the ICT impact comes from the channel of total factor productivity (TFP), because ICT are considered as a technology with broad application in several sectors of the economy, which finally increases TFP.

An empirical research has focused on several sectors of economic activity (e.g. agriculture, baking, railways), but the aggregate cross-country level is not so emphasized, despite the fact, that the measurement of aggregate technical efficiency and identifying the sources of technical inefficiency are important to promote economic growth. In this section some representative studies at the cross country level dealing with the impact of ICT are briefly described. Thompson and Garbacz [12] examined the impact of penetration rates of telecommunications on the technical efficiency in the ninety-three developed and developing countries during 1995-2003. Their results show important effects for low-income countries, but insignificant effects for OECD countries. Repkine [10] investigated whether telecommunications capital affects the level of technical efficiency in fifty developed and developing countries during 1980-2004 and found out that telecommunications capital positively affects the production efficiency in developing countries, but not in the developed ones. Winston and Chiang [14] analyzed the contribution of ICT in 25 countries over the period 1997-2006 and concluded that Eastern European countries gain more productive efficiency than the G7 countries when ICT are considered as a production factor. Dimelis and Papaioannou [3] examined the effects of ICT in reducing aggregate technical inefficien-

[1] VŠB-Technical University of Ostrava/Faculty of Economics, Department of Systems Engineering, Sokolská třída 33, 701 21 Ostrava 1, jan.mandak.st@vsb.cz.

cy and the impact of ICT on labour productivity across a panel of 42 developed and developing countries in 1993-2001 and confirmed both significant impact of ICT in reducing country inefficiencies and significantly positive ICT impact on labour productivity. Several studies at firm-level were also conducted (e.g. Becchetti et al. [2], Milana and Zeli [6]) and these studies concluded positive effect of ICT on firm-level technical efficiency.

In this study the technical efficiency at the aggregate cross-country level is estimated using a stochastic frontier analysis approach. Stochastic production frontier is simultaneously estimated with a technical inefficiency equation across a panel of 15 countries of the European Union during 1995-2007**.** The goal of this study is to examine the impact of ICT and non-ICT capital services on labour productivity, to find factors influencing technical inefficiency (with the emphasis on ICT) and to compare EU countries according to their efficiency in the period 1995-2007.

The structure of the paper is as follows: In the first section the theoretical background is briefly described. Section 2 presents an econometric specification of the production frontier and inefficiency equation. In the Section 3 economic indicators chosen as measures of output and inputs of production function and factors influencing inefficiency are characterized. In the 4th Section the obtained results of econometric estimation of production function and inefficiency equation are listed and in the last section the key results of the study are concluded.

2 Econometric specification

2.1 Production frontier and inefficiency equation

There are two approaches of estimation production frontier, prediction inefficiency levels and analysis of factors influencing inefficiency. In the two-step approach, production frontier is estimated and inefficiency levels are predicted, and, in the second stage, the inefficiency levels are regressed by a number of explanatory variables. According to Kumbhakar et al. [5], this two-stage estimation procedure has two drawbacks: inefficiency might be correlated with the inputs and the ordinary least squares method (OLS) does not take into account that the dependent variable of inefficiency takes only positive values. The second approach is a one-stage approach, which was first proposed in 1995 by Batesse and Coelli [1] and where the stochastic production frontier is simultaneously estimated with a technical inefficiency equation. Due to drawbacks of a two-stage approach mentioned above in this study the second one-stage approach is used. A Cobb-Douglas production function in the exponential form can be transformed into linear form using the natural logarithms and the production frontier, used for the analysis, takes the following form:

$$\ln LP_{it} = \beta_0 + \beta_1 \ln ICTH_{it} + \beta_2 \ln NICTH_{it} + (V_{it} - U_{it}). \tag{1}$$

In Equation (1), LP_{it} measures labour productivity (gross domestic product (GDP) per hour worked) for country i and year t, input factors of the production function are non-ICT capital services per hour worked $NICTH_{it}$ and ICT capital services per hour worked ICT_{it}. A Parameter β_0 is the intercept term, parameters β_1 and β_2 are output per hour worked elasticities of ICT and non-ICT capital. A random variable, which is independently and identically distributed with $N(0, \sigma_v^2)$, is denoted by V_{it}. The random error V_{it} should be independent of U_{it}, the nonnegative random error associated with the technical inefficiency of production. The random variable U_{it} has the half normal distribution, because the levels of efficiency can take only positive values.

The technical inefficiency effects are function of the set of explanatory variables z_{it} and the inefficiency equation is defined as:

$$U_{it} = z_{it}\delta + W_{it}, \tag{2}$$

where W_{it} is the random variable defined by truncation of the normal distribution while δs are parameters to be estimated. The technical efficiency of country i and year t can be obtained by taking:

$$TE_{it} = e^{-U_{it}}. \tag{3}$$

In this study, the technical inefficiency effects U_{it} are regressed by 4 explanatory variables:

$$U_{it} = \delta_0 + \delta_1 SHICT_{it} + \delta_2 EDUC_{it} + \delta_3 OPEN_{it} + \delta_4 EU12_{it} + W_{it}, \tag{4}$$

where $SHICT_{it}$ denotes share of ICT capital in total (*ICT + non-ICT*), $EDUC_{it}$ captures graduates (ISCED 5-6) in Maths, Science and Technology fields (as % of all fields), $OPEN_{it}$ stands for trade openness and $EU12_{it}$ is a dummy variable, which is equal to 1 if the country is in the EU-12 group and 0 otherwise. All these explanatory variables are described in detail in the next section.

The parameters in the stochastic production frontier model (1) and in the technical inefficiency Equation (2) as well as the total error variance and its two components $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and the parameter $\gamma = \sigma_u^2 / \sigma^2$ are estimated using maximum likelihood method. Using parameter γ it is possible to test, whether deviations from the frontier are due to inefficiency effects. If the null hypothesis that $\gamma = 0$ is rejected, than the inefficiency effects are a cause of the deviations from the production frontier

The statistical program *R* and its library *frontier* are used to obtain the parameters in the stochastic frontier Equation (1) and also the parameters in the inefficiency Equation (2).

3 Data and variables

Firstly, the data of the production function and secondly the data in the inefficiency equation are described. Output and inputs of the production function were obtained from the EU KLEMS database. The goal of the EU KLEMS Growth and Productivity Accounts is to support empirical and theoretical research of productivity in the European Union at the industry level. The big advantage of the EU KLEMS database is that it is based on analytical framework such as production functions and the theory of economic growth. This database includes measures of outputs (Gross Output, Gross Value Added) as well as various measures of inputs (Capital, Labour, Energy, Material or Service Inputs) since 1970.

Labour productivity for sector i and time t was selected as a measure of output and is defined as:

$$LP_{it} = \frac{GO_{it}}{H\_EMP_{it}}, \tag{5}$$

where GO_{it} is gross output at current basic prices (in millions of Euros) and H_EMP_{it} is total hours worked by persons engaged (millions). ICT capital services per hour worked (*ICTH*) and non-ICT capital services per hour worked (*NICTH*) in millions of Euros were chosen as measures of inputs. O´Mahony and Timmer [8] emphasize that in levels comparisons between countries purchasing power parities (PPPs) are needed to adjust output and inputs for differences in relative price levels. The values of PPPs were obtained from the database of the EU KLEMS project. A comprehensive description of the data contained in the EU KLEMS database is available in O´Mahony and Timmer [8].

In the second part of Section 3 the variables used as regressors in the inefficiency Equation (4) are identified. The first explanatory variable of the inefficiency equation is the share of ICT capital in total (*ICT + non-ICT*). Data for this variable are available in the EU KLEMS database. It is assumed, that countries with higher share of ICT capital should be more efficient than countries where the non-ICT capital prevails. The second explanatory variable represents the impact of education on inefficiency and is measured as graduates (ISCED 5-6) in Maths, Science and Technology fields (as % of all fields). Higher share of ICT capital should reduce country inefficiency, but it is also important to have qualified workers, who are able to use and understand modern technologies. The variables trade openness is measured as the share of imports plus exports in total GDP. International trade is considered as important channel of technology transfer, which affects the level of efficiency through higher competition. The last explanatory variable in the inefficiency is the dummy variable EU-12, which is equal to 1 if the country is in the EU-12 group and 0 otherwise. The group of EU12 countries consists of Belgium (BEL), Denmark (DNK), France (FRA), Germany (DEU), Greece (GRC), Ireland (IRL), Italy (ITA), Luxembourg (LUX), Netherlands (NLD), Portugal (PRT), Spain (ESP) and United Kingdom (GBR). These countries joined EU in the period 1 November 1993 - 31 December 1994 and it is expected that these countries are more efficient than the other ones. The data source of variables education, trade openness and EU12 is the Eurostat database. At the end of this Section we should emphasize that all explanatory variables are expected to have negative impact on technical inefficiency.

4 Discussion of results

4.1 Production frontier and inefficiency equation

Table 1 shows estimates of the stochastic production function (1) and also the inefficiency Equation (4) across a panel of 15 EU countries in 1995-2007. As it is evident, the elasticity of ICT capital is positive and highly significant which implies strong and positive impact of ICT capital on labour productivity. However, the impact of non-ICT capital is significantly negative. The study Dimelis and Papaioannou [3] also determined the negative impact of non-ICT capital on labour productivity. De Long and Summers [3] argue that the qualitative nature of other forms of capital, such as ICT, is more significant than the augmentation of fixed capital stocks, particularly in the case of foreign direct investments (FDI).

Before the estimation of parameters related to the sources of inefficiency, it should be tested, whether the deviations from the estimated frontier are due to inefficiency effects. From the Table 1 it is evident, that the parameter γ is significantly different from zero and we can proceed to the estimation. There was also no correlation between the explanatory variables which implies no multicollinearity problems. The second part of Table 1 contains results of the estimation of the technical inefficiency equation. A one percent rise in share of ICT capital results to the 0.012% decrease of inefficiencies among countries. Nevertheless, this variable is statistically significant at 27% level of significance. There are two variables, which are statistically significant at 90% level of confidence - education and dummy variable EU-12. The percent of graduates (ISCED 5-6) in Maths, Science and Technology fields is important for reducing country inefficiencies and implies the importance of technical education. The use of modern technologies supporting economic growth is not possible without qualified employees. Finally, the countries which are in the EU-12 group display significantly lower levels of inefficiency in comparison with countries, which are not members of the EU-12 group. The last variable in the inefficiency Equation (2), trade openness, didn't affect the technical inefficiencies, unlike in the study Dimelis and Papaioannou [4].

	Estimate	**Std. Error**	***p*-value**	
Production function				
Constant	3.282	0.309	< 0.0001	***
lnICTH	0.437	0.061	0.00001	***
lnNICTH	-0.038	0.018	0.03259	*
Inefficiency equation				
SH_ICT	-0.012	0.011	0.26831	
EU_12	-3.496	2.068	0.09086	.
EDUC	-0.042	0.025	0.08780	.
OPEN	0.001	0.008	0.93880	
σ^2	4.395	2.333	0.05956	.
γ	0.997	0.001	< 0.0001	***
log likelihood	-144.1909			

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 1 Maximum likelihood estimates for 15 EU countries (1995-2007)

4.2 Efficiency across countries

Table 2 presents average technical efficiencies for 15 EU countries in two time intervals 1995-2000 and 2001-2007 as well as in the entire period 1995-2007. Among ten most efficient countries there are eight countries of the EU-12 group and two Scandinavian countries Finland and Sweden, which are members of the EU-15 group (joined EU in 1 January 1995). Except of Belgium, in the group of five most inefficient countries there are not members of the EU-12 group. Twelve countries increase their efficiency from period 1995-200 to 2001-2007 and only in three countries (Slovenia, Czech Republic and Belgium) the efficiency slumped. Positive conclusion of the study is that an average efficiency for sample of 15 EU countries is each year higher than in the previous year (except of 1998 and 2006) and the annual growth rate is equal to 2.25 %.

Country	**1995-2000**	**2001-2007**	**1995-2007**	**Rank**
France	0.730	0.910	0.830	1
Italy	0.723	0.888	0.807	2
Netherlands	0.719	0.865	0.795	3
Finland	0.539	0.778	0.657	4
Germany	0.611	0.652	0.633	5
Sweden	0.505	0.742	0.621	6
Denmark	0.415	0.658	0.532	7
Spain	0.357	0.660	0.497	8
United Kingdom	0.318	0.574	0.437	9
Ireland	0.814	0.255	0.436	10
Slovenia	0.246	0.686	0.427	11
Hungary	0.240	0.641	0.407	12
Austria	0.275	0.416	0.343	13
Czech Republic	0.271	0.268	0.270	14
Belgium	0.875	0.092	0.260	15

Table 2 Estimated technical efficiencies

Table 3 confirms that the EU-12 group of countries is more efficient than the rest of countries in the sample. An average efficiency in the EU-12 group of countries is higher in the periods 1995-2000 and 2000-2007 as well as in the entire period 1995-2007.

Group	**1995-2000**	**2000-2007**	**1995-2007**
EU-12	0.618	0.617	0.581
non EU-12	0.346	0.589	0.454

Table 3 Comparison of EU-12 and non EU-12 countries

5 Conclusion

In this paper a stochastic production frontier approach was applied to simultaneously estimate stochastic Cobb-Douglas production frontier and technical inefficiency equation across a panel of 15 EU countries in the period 1995-2007. The results of the study provide strong evidence of positive significant impact of ICT capital on labour productivity. Impact of non-ICT capital on labour productivity was also significant, but negative. The share of ICT capital in total (*ICT + non-ICT*) reduce country inefficiencies, however is significant at the 26% level of significance. The study also found out the significance of technical education (measured as percent of graduates in Maths, Science and Technology fields) in reducing country inefficiencies. In comparison with not members of the EU-12 group, EU-12 countries achieved lower levels of inefficiency. The most efficient countries in the sample are France, Italy and Netherlands followed by the other EU-12 countries. Unlike in the study Dimelis and Papaioannou [4] trade openness didn't affect technical inefficiency of production in the EU. The obtained evidence suggest that ICT can be a powerful tool for development at the national level, but the employment of knowledge and technical workers is important for both the production and use of ICT.

Acknowledgements

This research was supported by Student Grant Competition of Faculty of Economics, VŠB–Technical University of Ostrava within project (SP2014/146).

References

[1] Battese, G. E., and Coelli, T. J.: A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics* **20** (1995), 325-332.

[2] Becchetti, L., Londono Bedoya, D. A., and Paganetto, L.: ICT Investment, Productivity and Efficiency: Evidence at Firm level using a Stochastic Frontier Approach. *Journal of Productivity Analysis* **20** (2003), 143-167.

[3] De Long, J. B., and Summers, L. H.: Equipment Investment and Economic Growth. *Quarterly Journal of Economics* **106** (1991), 445-502.

[4] Dimelis, S. P., and Papaioannou, S. K.: Technical Efficiency and the Role of ICT: A Comparison of Developed and Developing Countries. *Emerging Markets Finance and Trade* **47** (2011), 40-53.

[5] Kumbhakar, S., Ghosh, S., and McGuckin, J..: A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms. *Journal of Business and Economic Statistics* **9** (1991), 279-286.

[6] Milana, C., and Zeli, A.: The Contribution of ICT to Production Efficiency in Italy: Firm Level Evidence Using Data Envelopment Analysis and Econometric Estimations. *OECD Science, Technology and Industry Directorate*, Paris, 2002.

[7] OECD. The *economic impact of ICT: measurement, evidence and implications*. OECD, Paris, 2004.

[8] O´Mahony, M., and Timmer, M. P.: Output, Input and Productivity Measures at the Industry Level: The EU KLEMS Database**.** *The Economic Journal* **119** (2009), 374-403.

[9] Quiang, C. Z., Pitt, A., and Ayers, S.: Contribution of Information and Communication Technologies to Growth. World Bank, Washington, 2003.

[10] Repkine, A.: ICT penetration and aggregate production efficiency: empirical evidence for a cross section of fifty countries. *Journal of Applied Economic Sciences* **3** (2008), 65-72.

[11] Stiroh, K.: Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say? *American Economic Review* **92** (2002), 1559-1576.

[12] Thompson, H., and Garbacz, C..: Mobile, Fixed Line and Internet Service Effects on Global Productive Efficiency. *Information Economics and Policy* **19** (2007), 189-214.

[13] Van Ark, B., Inklaar, R., and McGuckin, R.: ICT and productivity in Europe and the United States: Where do the differences come from? *CESifo Economic Studies* **49** (2003), 295-318.

[14] Winston, T. L., and Chiang, Ch.: The impacts of country characteristics upon the value of information technology as measured by productive efficiency. *Production Economics* **132** (2011), 13-33.

Intervention in time series

Lubos Marek[1]

Abstract. The character of time series is often disturbed by external events that modify its course. It is events such as strikes, political changes, discoveries, new technologies, etc. These external events are called interventions. Intervention analysis was successfully used to solve many cases in which we failed with the classical approach. When we analyze time series with interventions, the first task is the estimation the value of intervention. Once we have estimated this value, we can either time series of interventions cleaned or intervention incorporated into the model. With such series we can able to work by common way – we can build the appropriate model and forecast the future values, etc.
In our article we describe the different types of interventions and their integration into the stochastic time series model. We work with transfer function models, which are mainly based on ARIMA models and on dynamic linear models. We describe the theory of interventions analysis including their detection in our article. This theory will be illustrated on practical data. Suitable example of intervention that had a significant impact on many macroeconomic time series, is the central bank's decision about the weakening of the CZK against the euro.

Keywords: time series analysis, stochastic process, interventions, interventions analysis.

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

The character of a time series is often affected by external events which change its time evolution. We have in mind here events such as strikes, political changes, discoveries of new technologies, etc. These events are called interventions. Time series containing interventions are processed with the aid of a technique called intervention analysis. Intervention analysis has successfully been used for solutions of many problems in which the traditional approach would not be adequate. The first task is to estimate the impact of an intervention. As soon as this impact is evaluated, we can either clean the time series, and then work with the cleaned time series in the usual way (for example, making predictions, which is often the main goal of time series analysis), or the intervention may be incorporated into the model. There is a question whether the impact of an intervention can really be evaluated in, say, an economic time series. Sometimes we are aware of a future intervention before it actually occurs; e.g., the action causing this intervention may have been planned a long time beforehand. In other instances the intervention is a real surprise for most of the public, not only the professional audience. Let us mention the weakening exchange rate of CZK with respect to EUR as a recent example. Consequences of this intervention are unmistakable in many macroeconomic time series, and its impacts are included in the topic of the present paper. We will not be much concerned with economic impacts but rather with the influence on the time evolution, which must be taken into account when analyzing the series. Only on the basis of that stage may economic consequences be examined. Transition-function models are used for modeling interventions; in other words, stochastic models of time series in combination with linear dynamic models are concerned. Details can be found, e.g., in [1], [2], [3], [5], and [8]. Statistical software package SCA was used for analyses – see [4] and [7] for more details.

2 Methodology

2.1 Two basic types of intervention variables

Let us introduce the two basic types of intervention variables now. The first type occurs at time T and is of a *temporary* character – it is only manifested at the given time T. This is called *pulse function*

[1] University of Economics, Prague/Department of Statistics and Probability, W. Churchill sq. 4, Prague 3, Czech Republic, marek@vse.cz.

$$P_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases} \tag{1}$$

Another type is *jump function*. It is an intervention at time T that has a *permanent* character; it is written in the form

$$S_t^{(T)} = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases} \tag{2}$$

The following relationship holds true between *jumps* and *pulses* (as they are sometimes referred to):

$$P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)} = (1 - B) S_t^{(T)} \tag{3}$$

Hence it is clear that an intervention model can solely be expressed by either a jump or a pulse function. The jump function is described by the following formula:

$$S_{t+j}^{(T)} = S_t^{(T-j)} \qquad j = ...,-1,0,1,... \tag{4}$$

2.2 Intervention and transition-function model

Let us now have a closer look at individual interventions. Before that, we have to point out that intervention models represent a special case of transition-function models. This aspect should be apparent from the definitions of *jump* and *pulse functions*. Replacing symbols S and P with X, we can view a jump or pulse function as a special input series in the transition-function model, or a dynamic regression model. The input series has a specific form in this case, namely, it is a deterministic sequence consisting of zeros and ones. A general form of the transition-function model is given as follows

$$Y_t = c + \frac{\omega_h(B) B^k}{\delta_r(B)} X_t + N_t \tag{5}$$

where

$$\omega_h(B) = \omega_0 + \omega_1 B + ... + \omega_h B^h \tag{6}$$

and

$$\delta_r(B) = 1 - \delta_1 B - ... - \delta_r B^r \tag{7}$$

N_t is *noise series*, and B is the classical back-shift operator.

2.3 Pulse interventions

Let us once more emphasize that a pulse intervention has a *temporary character*. Re-writing (1) for variable X, we get

$$X_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases} \tag{8}$$

In other words, $X_t^{(T)}$ is therefore a binary variable, which could be understood as a switch (0 meaning "off" and 1 meaning "on"). Because a pulse intervention does not have a permanent character, it can be manifested in different ways; alternatively, we can say that this intervention has *different responses*.

2.3.1 One-point response

A typical one-point, or one-period, response is illustrated in Figure 1, which shows a time series with a stationary mean value. Apart from a pulse intervention at time T, time series Y_t is constant. At the said time the value of Y_t is shifted up or down with respect to the previous constant level, but immediately after time T it goes back to the original level. In this case the response really is one-point and the transition function is simply expressed as $\omega_0 X_t$ where ω_0 is the size of the shift at time T.

The situation is rather different when the mean value is non-stationary, governed by a linear trend. The response is the same, but the values of the series after the intervention still follow the linear trend. Figure 2 shows the latter case. The transition function's form is the same as in the previous instance, i.e., $\omega_0 X_t$ where ω_0 is the size of the shift at time T.

2.3.2 Many-point response

A pulse intervention may bring about a many-point response, which in its turn may be either *dynamic* or *time-dispersed*. This means that the pulse intervention's response occurs not only at time T but also at additional moments occurring after T. In Figure 3 we can see the evolution of a series that is constant except for the intervention effect. Figure 3 is nearly the same as Figure 1, the only difference being that series Y_t responds to the intervention also at time $T+1$. The transition function's form is in this instance, namely, $(\omega_0 + \omega_1 B) X_t$.. Let us take, for the sake of an example, $\omega_0 = 200$ and $\omega_1 = 250$. Then the value Y_t is increased by *200* units at time T and by *250* units at time $T+1$. However, these effects are not cumulative, that is, they do not sum up. Since it is $X_t = 0$ after time $t = T$, of course it is $\omega_0 X_t = 0$ after time $t = T$; as well as $\omega_1 B X_t = 0$ after time $t = T$. Hence, if the original value of Y_t is increased by *200* units at time T, such increase does not occur after time T. Similarly, the increase at time $T+1$ is only by *250* units, not by *200+250* units. And after time $T+1$, there is no increase at all (i.e., the value of the increase is *0* units).

Figure 4 shows a continuous dynamic response following after time T. In this case each consecutive response after time T is a fraction of the preceding one because in each step it is divided by the value of parameter δ_1. The chart in Figure 4 has the same shape as that of the Koyck model, which has the same type of the transition function. Namely, the form of the transition function is $[\omega_0 / (1-\delta_1 B)] X_t$, i.e., it really corresponds to the Koyck model. Parameter ω_0 is the response's initial value, and δ_1 determines the rate of decrease for $t > T$. Let us set $\omega_0 = 200$ and $\delta_1 = 0.10$. The following Table shows the response values at times following after T:

t	Response
T	$\omega_0 = 200$
$T+1$	$\delta_1 \omega_0 = 0,5 \cdot 200 = 100$
$T+2$	$\delta_1^2 \omega_0 = 0,5^2 \cdot 200 = 50$
…	…
$T+k$	$\delta_1^k \omega_0 = 0,5^k \cdot 200$

Table 1 Response values

In this example, the value of coefficient δ_1 complies with the stability condition due to $|\delta_1| < 1$. If the value of k went to infinity, the limit of the response would be zero because the value of δ_1^k goes to zero for k going to infinity. As already mentioned above, the responses do not sum up. After time $t = T$ we only have responses dispersed in time, and their values depend on the initial value of ω_0 and the value of coefficient δ_1. The permanent increment g can be determined as the total cumulated response for the relevant transition function. Setting $B = 1$ in the transition function, we get $g = \omega_0 / (1-\delta_1)$ in our example, resulting in 200/0.5=400. It is problematic to interpret this increment because g is defined as the overall change in Y_t if X_t continually grows by the unit value. The value X_t, however, grows from zero to one only once, i.e., temporarily at time T.

Figure 5 is similar to Figure 2, which showed a series with a non-stationary mean value. Evolution in the chart is different by the dynamic response of Y_t to the dynamic intervention following the response at time T: the response gets gradually weaker at $T+1$, $T+2$, …. We have thus come to the known Koyck model again. The transition function has the form $[\omega_0 / (1-\delta_1 B)] X_t$; and series Y_t and X_t will have to be differenced to remove the non-stationarity. Again $|\delta_1| < 1$ holds true. With the weakening response in Y_t, the values go back to the level implied by the series' non-stationary character.

Figure 6 finally shows a special case of the Koyck-type pulse intervention with parameter $\delta_1 = 1$. Now the pulse intervention looks more like a jump because the change has a permanent character here. The following explanation shows why this intervention is classified as a pulse: parameter δ_1 determines the rate of intervention's fading out in the series Y_t after time T (of course, with the initial value of ω_0). If, for example, there is $\delta_1 = 0.5$, each consecutive response is δ_1-times smaller than the preceding one; 0.5-times smaller in our case. If there is $\delta_1 = 1$, each consecutive response is δ_1-times smaller than the preceding one – 1-times smaller in this case – so that the response does not change and remains on the same level after time T. The latter case is a rather special one; the same effect can be achieved with the aid of a jump intervention, as we will see below.

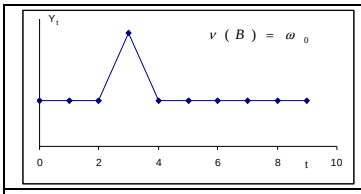


Figure 1 Pulse – type 1

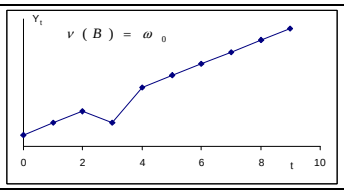


Figure 2 Pulse – type 2

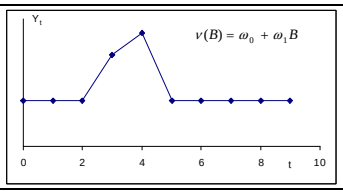


Figure 3 Pulse – type 3

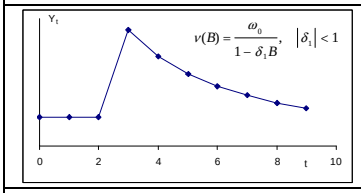


Figure 4 Pulse – type 4

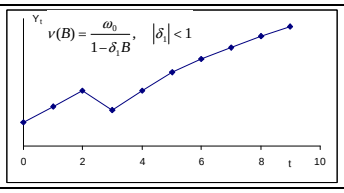


Figure 5 Pulse – type 5

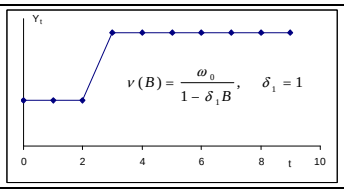


Figure 6 Pulse – type 6

2.4 Jump interventions

Let us once more emphasize that a jump intervention has a *permanent character*. Re-writing (2) for variable X, we get

$$X_t^{(T)} = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases} \tag{9}$$

In other words, $X_t^{(T)}$ is a binary variable, which could again be understood as a switch (0 meaning "off" and 1 meaning "on"). A jump intervention has a permanent character and may be manifested in different ways, i.e., the responses may be different.

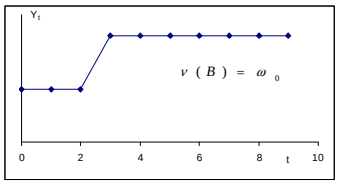


Figure 7 Jump – type 1

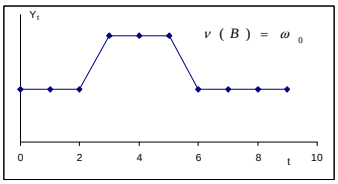


Figure 8 Jump – type 2

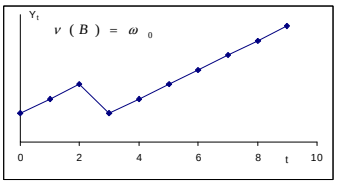


Figure 9 Jump – type 3

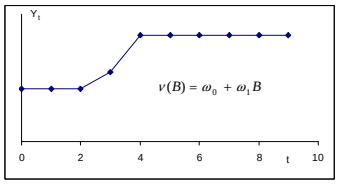


Figure 10 Jump – type 4

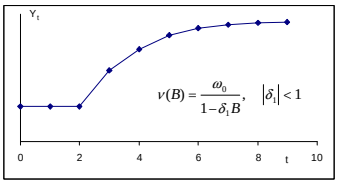


Figure 11 Jump – type 5

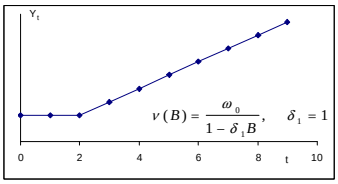


Figure 12 Jump – type 6

2.4.1 One-point permanent response

Figure 7 shows a typical evolution of a jump intervention. The stationary series Y_t is affected by an intervention at time T; the level of the series, constant to that moment, is shifted upwards (in this particular case) or downwards. The series thus gets to a new level and remains there. The transition function of the respective dynamic regression model is $\omega_0 X_t$ where ω_0 is the size of the shift at time T. If the parameter value is $\omega_0 = 100$, the value of Y_t is permanently increased by 100 units; conversely, if $\omega_0 = -100$ then Y_t shows a permanent decease by 100 units.

Figure 8 shows an intervention whose character is slightly different. The jump intervention variable in this case behaves like an identical pulse in three consecutive moments of time and can be written as follows:

$$X_t^{(T)} = \begin{cases} 0, & t < T,\ t > T+2 \\ 1, & t = T, T+1, T+2 \end{cases} \tag{10}$$

Figure 9 shows a jump intervention in a non-stationary series. This series has an apparent linear trend (consequently, its mean value is non-stationary). It linearly grows until time T, when the intervention causes a

decrease by a value of ω_0; and after time T the series grows with a linear trend again, but on a *permanently lower level*. This is a difference compared with the situation of Figure 2, in which a pulse intervention caused a decrease of the series at time T, and then the series grew with a linear trend *on the original level*. In Figures 10 through 12 we can see other instances of jump interventions. Their detailed descriptions can be found, e.g., in [6].

2.5 Example

The following Figure shows the daily time series of the CZK/EUR exchange rate. The source of the date is the Czech National Bank (CNB – www.cnb.cz) and the series covers the time period from the beginning of 2013 until April 12, 2014. The chart clearly shows the CNB intervention at the beginning of November 2013, when the CNB made a decision to artificially weaken the CZK exchange rate with respect to EUR, which was manifested by a large jump divided into several consecutive days.

Figure 13 CZK vs EUR Rate

At first sight it is clear that it is a jump function with a many-point response – cf. Figure 10. We now have to estimate a suitable model for this time series and incorporate the jump intervention into the model. After a number of analyses (study of stationarity, ACF, PACF, EACF, and other identification methods) the following model was identified as the most suitable one:

$$(1-\theta_1)Y_t = (\omega_0 + \omega_1 B) X_t + \varepsilon_t \tag{11}$$

where Y_t is the CZK/EUR exchange rate time series, X_t is an intervention variable, and ε_t is normally distributed white noise. Intervention variable X_t is binary, defined as

$$X_t^{(T)} = \begin{cases} 0, & t \le 216 \\ 1, & t > 216 \end{cases} \tag{12}$$

Its specific form is

$$Y_t = 0.9801 Y_{t-1} + 0.9626 X_t - 0.9595 X_{t-1} + \varepsilon_t . \tag{13}$$

```
SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- INTERV
-------------------------------------------------------------------------
VARIABLE   TYPE OF    ORIGINAL     DIFFERENCING
          VARIABLE OR CENTERED
  RATE     RANDOM     ORIGINAL     NONE
  SHIFT    BINARY     ORIGINAL     NONE
-------------------------------------------------------------------------
PARAMETER   VARIABLE  NUM./  FACTOR  ORDER   CONS-     VALUE      STD     T
  LABEL       NAME    DENOM.                TRAINT              ERROR  VALUE
  1    W0     SHIFT    NUM.      1      0     NONE      .9626     .0628  16.92
  2    W1     SHIFT    NUM.      1      1     NONE     -.9595     .0630 -16.83
  3   TH1      RATE     AR       1      1     NONE      .9801     .00026 24.03

EFFECTIVE NUMBER OF OBSERVATIONS . .           323
R-SQUARE . . . . . .                          .994
RESIDUAL STANDARD ERROR. . . . . .   .626679E-01
```

Figure 14 CZK vs. EUR Rate

The above-described intervention occurred at time $T = 217$. This model was successfully confirmed at several stages of verification – tests of stationarity and of unit roots, as well as analyses of residua of the input and output series. The quality of this model is very good, with a determination index value of 0.994 – cf. the output. The series Y_t itself is governed by an AR(1) model and the value of parameter θ_1 is close to one, as could be expected (this model is close to a random walk).
The quality of this model is also clear from the autocorrelation and partial autocorrelation functions, whose values do not significantly depart from zero.

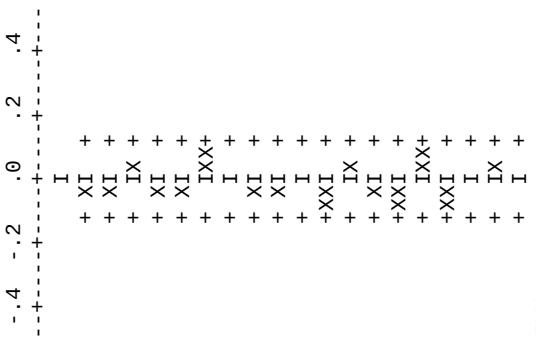


Figure 15 ACF

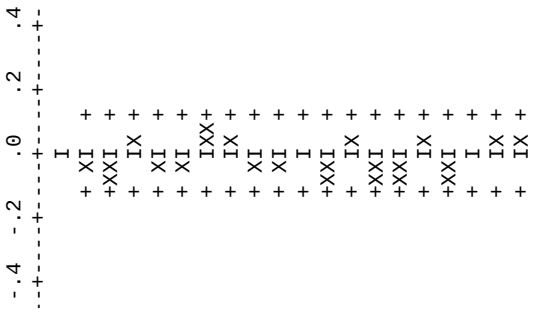


Figure 16 PACF

3 Conclusions

Interventions significantly affect the evolution of a time series. We illustrate this fact in an example describing the intervention by the Czech National Bank. It is clear that the intervention should be estimated and incorporated into the model of the time series. The time series itself and its error component are governed by a stochastic process (ARIMA or SARIMA); hence the classical regression analysis cannot be utilized for their modeling. More sophisticated techniques have to be used – dynamic regression models and, in particular, transition-function models. Here the intervention variable plays the role of the input series. The model we have created can be used for subsequent analysis or for making predictions. The model quality is very good, its determination index is close to one, and the model successfully passed all stages of verifications. This model will be further used in analyses to follow this paper.

References

[1] Abraham, B. – Ledolter, J.: Statistical Methods for Forecasting, John Wiley & Sons, New York, 1983.
[2] Anderson, O. D.: Time series analysis and forecasting – Box-Jenkins approach. London, Butterworth 1976.
[3] Box, G. E. P. – Jenkins, G. M. – Reinsel, G.C.: Time series analysis, forecasting and control. Third edition. Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1994.
[4] Forecasting and Time Series Analysis using the SCA Statistical System. Volume 1. Scientific Associates Corp. Oak Brook, Illinois, USA, 1991.
[5] Granger, C. W. J., Newbold, P.: Forecasting Economic Time Series. Academic Press. New York, 1986.
[6] Pankratz, Alan: Forecasting with dynamic regression models. John Wiley & Sons inc., New York, 1991.
[7] The SCA Statistical System. Reference manual for general statistical analysis. Scientific Associates Corp. Oak Brook, Illinois, USA, 1991.
[8] Wei, W. W. S.: Time series analysis – Univariate and multivariate methods. Redwood City, California, Addison-Wesley Publishing Company 1990.
[9] ic Associates Corp. Oak Brook, Illinois, USA, 1991.
[10] Wei, W. W. S.: Time series analysis – Univariate and multivariate methods. Redwood City, California, Addison-Wesley Publishing Company 1990.

A New Consistency Measure for Additive Preference Matrix in Pairwise Comparisons

Markéta Matulová[1]

Abstract. A lot of methods for solving multiple criteria decision problems require some weights of criteria. Widely used way of setting weights is based on pairwise comparisons, but there often occur inconsistencies in the preference expression. The article focuses on properties of pairwise comparison matrix, that influence the possibility to obtain relevant weights. There is suggested a new consistency index for the additive preference matrix and its performance is compared with traditional indices in a numerical study.

Keywords: multicriteria decision making, analytic hierarchy process, pairwise comparisons, additive preference matrix, consistency.

JEL Classification: M10, M20
AMS Classification: 90B50, 91B06

1 Introduction

One of the crucial steps in most of multicriteria methods is to quantify the criteria relevance. It is often expressed in a relative way by means of weights $w_1, \ldots, w_n$ assigned to the n decision criteria by some process of pairwise comparisons. This procedure was originally designed for the analytic hierarchy process (AHP), where it is also used for determining the ranking of the alternatives. Preferences between compared elements (criteria or alternatives) can be expressed in different ways. They are usually arranged in a *pairwise comparison matrix* (PCM).

In this article, there are examined properties of pairwise comparison matrices. The text is structured as follows. The first part deals with multiplicative preference information and the way of expressing the degree of its consistency. We give an overview of existing consistency indices and mention their properties. The next part is oriented on the additive approach. We suggest a new consistency index for the additive preference matrix as a solution to an optimization problem. The relevance of the index is discussed theoretically. As a quadratic programming problem it can be more computationally convenient than the other approaches. Empirical comparison of the new index and index designed by Ramík [7] is given in the final part. According to the numerical results, the performance of our index is equivalent to that of traditional index. In the conclusion we consider the advantages of the additive approach and suggest the direction for further research.

2 Multiplicative preferences

According to widely used Saaty's method [8], the preference information is expressed by intensities from a given scale $\{1/9, 1/8, \ldots, 1, \ldots, 8, 9\}$, where the value of 1 represents indiference between options and the value of 9 represents absolute preference of one option to the other (and $1/9$ vice versa), other values are intermediate. The intensities are arranged in a pairwise comparison matrix A, where the entries to this matrix a_{ij} are supposed to estimate the ratios w_i/w_j (so the reciprocity condition $a_{ij} = 1/a_{ji}$ is required). The expert in charge of assessing should be coherent in expressing his opinion. The elements of fully consistent matrix A fulfill the condition

$$a_{ij} \cdot a_{jk} = a_{ik} \quad \text{for all triplets } i, j, k \tag{1}$$

[1] Faculty of Economics and Administration, Masaryk University, Lipová 41a, Brno, the Czech Republic, 8987@mail.muni.cz.

Unfortunately, it is almost impossible to be absolutely consistent, especially if there are a lot of compared elements. The degree of consistency can be expressed by a measure of consistency, the traditional Saaty's consistency index $CI(A)$ is defined as

$$CI(A) = \frac{\lambda_{max}(A) - n}{n - 1}, \tag{2}$$

where $\lambda_{max}(A)$ is the principal eigenvalue of the matrix A. Despite its name the consistency index expresses the degree of inconsistency. Characterization of the index is given in [9] and some of its properties are described in [5]. It can be proved that the totally consistent matrix A has rank 1 and its only nonzero eigenvalue $\lambda_{max}(A)$ is equal to n with corresponding eigenvector $w = (w_1, \dots, w_n)$ being the weight vector (then index $CI(A) = 0$). If the entries a_{ij} change slightly, the eigenvalues change too. The principal eigenvalue $\lambda_{max}(A)$ will be slightly greater than n while the others will remain close to zero, so $CI(A)$ is reasonably defined. Some authors [5] estimate the principal eigenvector w by normalization of geometric means of the rows of A and principle eigenvalue by multiplying the sum of the columns of A by the vector w.

In order to eliminate the influence of dimensionality, the consistency ratio $CR(A)$ is introduced by the formula

$$CR(A) = \frac{CI(A)}{RI(n)}, \tag{3}$$

where $RI(n)$ is the average value of CI for randomly generated pairwise comparison matrices of the same size n. For determining the value of average consistency index different authors used different simulation methods and their results are more or less close, see Alonso and Lamata's overview [1]. According to Saaty a matrix A should be accepted as sufficiently consistent if its consistency ratio is small enough: $CR(A) < 0.1$. If the condition is not fulfilled, then a revision of the judgements is recommended.

2.1 Alternative consistency indices

Barzilai criticizes Saaty's approach in [2]. According to him, characteristics $CI(A)$ and $CR(A)$ are heuristics with poorly understood properties. Moreover, he claims that the statement "the closer $\lambda_{max}(A)$ to n, the more consistent matrix" is not justified anywhere in the AHP literature. But this is not true anymore: although Saaty didn't propose his index $CI(A)$ involving any optimization process, Fedrizzi [6] describes a metric that is minimized by the consistent matrix where weight vector is determined as the principal eigenvector. The distance of A from the nearest consistent matrix in this metric is exactly $CI(A)$. Fedrizzi studies geometrical characterization of inconsistency evaluation generally in [6] and defines inconsistency index of matrix A as its distance from the set of consistent matrices of the same size. He shows that the usage of different metrics leads to different indexes, for example usual Euclidean distance induces the Least Square Index LS defined by Chu,

$$LS = min_{W_1,\dots,W_n} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left(a_{ij} - \frac{w_i}{w_j}\right)^2, \qquad s.t. \sum_{i=1}^{n} w_i = 1, w_i > 0 \tag{4}$$

or the logarithmic distance induces the Geometric Consistency Index GCI defined by Crawford and Williams,

$$GCI = \frac{2}{(n-1)(n-2)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ln^2 \left(a_{ij} \frac{w_j}{w_i}\right) \tag{5}$$

The logarithmic approach is very useful, because by applying a logarithmic function to all consistent matrices componentwise we get a set of matrices that forms a linear subspace in $R^{n\times n}$. So, after the logarithmic transformation, the consistent approximation C of the given matrix A is its orthogonal projection onto this subspace. Barzilai [2] suggests to decompose matrix A to its consistent component C and the error matrix E. This decomposition can be used instead of distorting the answers by forcing the values of judgement to improve consistency.

Peláez and Lamata defined another promising index

$$CI^* = \frac{1}{\binom{n}{3}} \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \left(\frac{a_{ik}}{a_{ij}a_{jk}} + \frac{a_{ij}a_{jk}}{a_{ik}} - 2 \right) \tag{6}$$

In [3] there are described some other consistency indices, namely

- Relative Error; Barzilai – 1998
- Shiraishi, Obata, Daigo – 1998
- Harmonic Consistency Index (HCI); Stein, Mizzi – 2007
- Cavallo, D'Apuzzo – 2009
- Ramík, Korviny – 2010

2.2 The properties of the indices

Fedrizzi and Brunelli [4] stated five axioms aimed at characterizing consistency indices and they have shown that the system of axioms is logically consistent and that the axioms are independent.

1. (A1) Existence of a unique element representing consistency
2. (A2) Invariance under permutation of alternatives/criteria
3. (A3) Monotonicity under reciprocity-preserving mapping
4. (A4) Monotonicity on single comparisons
5. (A5) Continuity

The first two axioms are just regularity conditions, the last one guarantees that small change in the input generates only small change in the value of index. Axioms (A3) and (A4) represent the requirement that the index does not decrease when the matrix moves away from consistency. According to [4], Saaty's index CI defined by (2), Crawford and Williams index GCI defined by (5) and Peláez and Lamata index CI^* defined by (6) satisfy all of the presented axioms. The most demanding axioms seem to be monotonicity requirements (A3) and (A4), other above mentioned indices violate at least one of them.

3 Additive preferences

Alternative approach to expressing the preference intensity is the use of the (nonnegative) additively reciprocal preference matrix B, so that $b_{ij} + b_{ji} = 1$, for every $i, j \in \{1, \ldots, n\}$. This approach should be more convenient to carry out. The assessing expert assigns the values $b_{ij} = b_{ji} = 0.5$ if the elements i, j are equally preferred or assigns $b_{ij} = 1$ and $b_{ji} = 0$ if he absolutely preferes i-th option to the j-th one, or he can use something in between: he always divides the unit into two parts according to the preference of the compared elements (for practical reasons it is reasonable to restrict the range of b_{ij} to the interval $(0,1)$, the extreme values are usually set to 0.01 and 0.99) . One possibility how to set the weights $w_1, \ldots, w_n$ is the solution of the system of equations

$$b_{ij} = \frac{w_i}{w_i + w_j}, \qquad i, j \in \{1, \ldots, n\} \tag{5}$$

The solution exists under the condition, that the matrix B has the multiplicative-transitivity property $b_{ij}.b_{jk}.b_{ki} = b_{ik}.b_{kj}.b_{ji}$ for all triplets of i, j, k [7]. If this condition is violated, there is no solution to the system (5) and the weights have to be approximated.

Let us suppose that B is positive additively reciprocal matrix. Ramík [7] suggests to evaluate the consistency of the matrix B by the additive consistency index $CI_a(B)$ as follows

$$CI_a(B) = CI(B^{mod}), \tag{6}$$

where the matrix B^{mod} consists of elements $u(b_{ij})$, $i, j \in \{1, \ldots, n\}$ transformed by a function u defined as

$$u(x) = \frac{x}{1 - x} \tag{7}$$

The consistency index CI_a defined by (6) can be used for a construction of a rule for accepting matrix as sufficiently consistent. Ramík [7] defined a consistency ratio similarly to the CR introduced in (3). We will modify this idea a bit.

3.1 A new approach to constructing additive consistency ratio

Original Saaty's idea of multiplicative consistency ratio was to evaluate consistency of a given matrix A in comparison to a matrix of the same size n, where the preferences were chosen randomly. So we can consider additive preference matrix B in context of randomly constructed additive preference matrices. Let us generate a set of random positive additively reciprocal matrices and compute their consistency indexes. From 500000 runs of such simulation for matrices of size $n \in \{3,..., 10\}$ we compute values of average consistency index $RI_a(n)$, and compare them to Alonso and Lamata's [1] values of multiplicative random indices $RI(n)$ in Table 1. All computations were performed in Matlab.

n	3	4	5	6	7	8	9	10
$RI(n)$	0.5245	0.8815	1.1086	1.2479	1.3417	1.4056	1.4499	1.4854
$RI_a(n)$	0.5642	0.9833	1.2899	1.5175	1.693	1.831	1.943	2.032

Table 1 Average additive consistency index $RI_a(n)$ vs. multiplicative index $RI(n)$

For a matrix B of size n we propose following definition of additive consistency ratio $CR_a(B)$,

$$CR_a(B) = \frac{CI_a(B)}{RI_a(n)} \tag{8}$$

It is recommended to accept matrix B as consistent enough if $CR_a(B) < 0.1$.

From Table 1 we can see that values of additive random indexes $RI_a(n)$ are slightly larger those of $RI(n)$. There is an explanation for this fact. Let us suppose that the entries b_{ij} of the upper triangle of the matrix B have the uniform distribution over the interval $(0, 1)$. In the computation of additive consistency index, there is involved matrix B^{mod} whose elements are $u(b_{ij})$ and their range is not restricted. It can be proved that their cummulative distribution function $F(x)$ is the inverse of the transformation function $u(x)$ defined by (9),

$$F(x) = u^{-1}(x) = \frac{x}{x+1}, \quad x \geq 0 \tag{9}$$

For practical computations of the elements of B in Table 1 we restricted the interval of their distribution to (0.01,0.99) in order to avoid generating "close to zero" elements and subsequent "devision by zero" problems in transformation of B.

3.2 A new consistency index and its properties

Another possibility how to measure inconsistency of the additive PCM B is to express how much is the system (5) violated. If we rewrite the equations in (5) using the additive reciprocity, we get equivalent conditions

$$b_{ij}w_j - b_{ji}w_i = 0, \quad i, j \in \{1, \dots, n\} \tag{10}$$

Assuming that weights are normalized to sum to one, we can define least squares index

$$LSI(B) = min_{w_1,\dots,w_n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(b_{ij}w_j - b_{ji}w_i\right)^2, \quad s.t. \quad \sum_{i=1}^{n} w_i = 1, w_i > 0 \tag{11}$$

The computation of the index value is based on the solution of the quadratic programming problem, so it is easy to perform. The index can be standardized by a number of elements in the sum, $n(n-1)/2$.

Let us examine the properties of the additive indices in the context of axioms stated by Fedrizzi and Brunelli. The reformulation of the axiom (A3) for the additive case is not straightforward, so we have concetrated on the other four axioms. Ramík's index $CI_a(B)$ satisfies these conditions, because it is the Saaty's index applied on the transformation $u(B)$ of the PCM B, where the transformation u defined by (7) is continuous and increasing on the interval (0,1). The properties of the new index LSI are similar. It can be shown that for all additively consistent matrices with nonzero elements there exists a positive weight vector w, for which the condition (5) holds. Dividing such a weight vector by a sum of its elements we get a solution to problem (11) with

zero value of the objective function, so every additively consistent matrix B satisfies $LSI(B) = 0$. The property (A2) is clearly satisfied because of the symmetry in the objective function and constraints in the optimization problem (11). The continuity property (A5) is known from sensitivity analysis in quadratic programming. To prove the monotonicity still remains an open problem.

3.3 Numerical comparison of the indices

We have computed the values of the indices CI_a and LSI for 1000 randomly generated positive additively reciprocal matrices of different sizes $n \in \{3, \dots 10\}$ and searched for the level of their agreement. It can be expressed graphically, see the Figure 1, where is a scatterplot for $n = 6$.

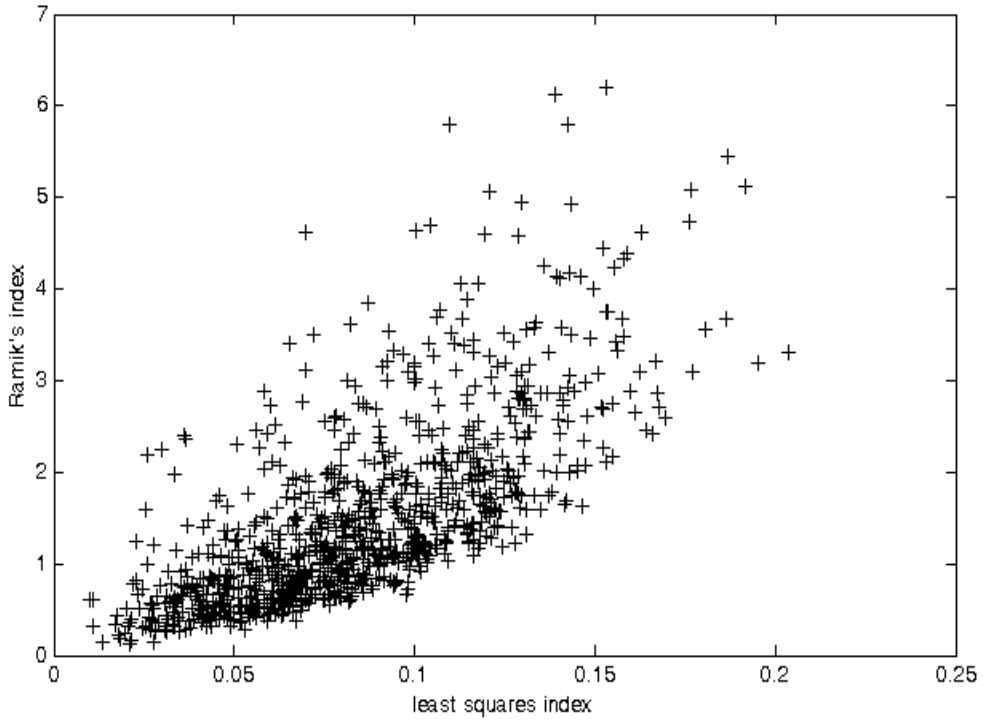


Figure 1 Scatterplot of the $CI_a(B)$ and $LSI(B)$ for randomly generated PCMs B

As we can see in the Figure 1, there is a good accordance in the performance of the indices. The linear correlation between the indices expressed by Pearson coeficient ρ and the rank correlation expressed by Spearman coeficient ρ_s are computed in Table 2.

n	3	4	5	6	7	8	9	10
ρ	0.81	0.75	0.73	0.67	0.66	0.66	0.65	0.64
ρ_s	0.90	0.80	0.76	0.72	0.70	0.70	0.68	0.66

Table 2 Correlation coefficients ρ and ρ_s of $CI_a(B)$ and $LSI(B)$ for 1000 randomly generated PCMs B of size n

4 Conclusion

We examined different consistency indices for the multiplicative preference matrices and their properties in order to suggest reasonable measure of consistency in the additive case. We studied Ramík's additive index CI_a and proposed a new index *LSI*. Their properties were studied theoretically and their performance was compared

empirically in a simulation study. Both indices fulfill the regularity conditions (they give a zero value for all consistent matrices and are invariant to the permutation of elements) and are continuous functions of their arguments (the entries of PCM). The monotonicity condition should be further examined in some other study. As we can see from numerical results, the indices perform similarly.

A great advantage of additive PCMs is the simplicity of the preference expression. In the future we intend to study the possibility of expressing the expert's opinion in the form of „sufficiently consistent matrix" from a psychological point of view. The traditional Saaty's scale with linguistic equivalents is simple too, but as we already know, the restricted scale is one of the main reasons of inconsistent answers. On the other hand, the extension of the scale would be more complicated for the expert to express his opinion verbally. The additive interpretation of dividing 100 percent into two alternatives is hopefully a solution to this problem.

References

[1] Alonso, J. A., Lamata, M. T.: Consistency in the analytic hierarchy process: a new approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **18** (2006), 445 – 459.
[2] Barzilai, J.: Consistency Measures for Pairwise Comparison Matrices. *Journal of Multi-criteria Decision Analysis* **7** (1998), 123 – 132.
[3] Brunelli, M., Canal, L. Fedrizzi, M.: Inconsistency indices for pairwise comparison matrices: a numerical study. *Annals of Operational Research* **145** (2013), 493-509.
[4] Brunelli, M., Fedrizzi, M.: Axiomatic properties of inconsistency indices for pairwise comparisons. *Journal of the Operational Research Society*, 2013.
[5] Chen, Q., Triantaphyllou, E.: Estimating Data for Multi-Criteria Decision Making Problems: Optimization Techniques. *Encyklopedia of Optimization*, (2001), Kluwer Academic Publishers, Boston, 27-36.
[6] Fedrizzi, M.: Distance-Based Characterization of Inconsistency in Pairwise Comparisons. *Advances in Computational Intelligence*, 2012, 30-36.
[7] Ramík, J.: Measuring transitivity of fuzzy pairwise comparison matrix. *Proceedings of 30th International Conference Mathematical Methods in Economics* (Ramík, J. and Stavárek, D., eds.). Silesian University in Opava, School of Business Administration, Karviná, 2012, 751-756.
[8] Saaty, T. L., The Analytic Hierarchy Process. McGraw-Hill,New York, 1980.
[9] Saaty, T. L., Decision making with the analytic hierarchy process, *Int. J. Services Sciences* **1** (2008), 83-98.

Estimating Capital Requirement According to Solvency II and its Impact on Insurance Companies

Petra Matušková

VŠB – TU Ostrava, Sokolskátřída 33, Ostrava, 701 21,
Czech Republic, e-mail: petra.matuskova@vsb.cz

Abstract. Economic situation which has been on the financial markets in the last few years has led to the changes in the regulation of financial institutions. Insurance market is the last area which is awaiting key form of regulation; it should become effective in 2016 when the long prepared regulatory regulations Solvency II will come into force. This will affect not only regulatory practice but also insurance companies' operation itself and that especially in the area of risk, capital management and information system. Main aim of this paper is to estimate capital requirements for currency risk by utilizing various methods of Monte Carlo simulation; compare acquired results and evaluate the impact of new regulatory approach on the insurance companies. The portfolio of insurance company will consist of selected stocks indexes and the trend of their yields will be simulated by Monte Carlo simulation, Antithetic Sampling Monte Carlo and Latin Hypercube Sampling Monte Carlo with dependence. The method Value at Risk will be used to determine capital requirements. This method is legislatively recognized by directive Solvency II.

Keywords: Monte Carlo simulation, Solvency II, capital requirements, insurance
JEL Classification: G17, G32
AMS Classification: 65C05, 60J65

1 Introduction

Insurance sector represents one of the most complex sectors of financial services which aim is to eliminate negative impact of unpredictable events. The current time of political and economic instability and also the development of European market lead to the fact that the emphasis is placed on credibility, transparency and the stability of financial institutions. In order to ensure the client's protection and all of the other aforementioned aspects it is necessary to determine certain rules for the entrepreneurship in this area including the establishment of supervisory institutions. Important role is played by the risk management process which includes identification, control, and financial risk coverage without which no financial institution can operate and therefore more attention is paid to this area.

Since the 70s of last century the regulation and supervision of the insurance market has been continuously harmonized by the implementation of three generation of directives and consequently by Solvency I regime. This legal solvency regime and the regulation is insufficient and it does not enable to detect all risks and it also does not require that the insurer takes into consideration the environment of low interest rates. Therefore it has been substituted in 2009 by legally amended Solvency II[1] which is based on the regulation of risk management, balance approach and it implements more fundamental and more complex evaluation of a total financial situation of the insurer. Solvency II also provides the basic tool for risk management and is based on three pillars (requirements for sufficient capital of the insurance companies, rules for risk management and requirement for the insurance's activity transparency); it creates the preconditions for fulfilling basic philosophy of EU, meaning to protect the consumer in market conditions. Full implementation of Solvency II is assumed in 2016, see EIOPA [8].

[1]More detailed information pertaining to the Solvency II problematic can be found at web pages of the European Insurance and Occupational Pensions Authority

Main aim of this paper is to estimate capital requirements for currency risk by utilizing various methods of Monte Carlo simulation; compare acquired results and evaluate the impact of new regulatory approach on the insurance companies.

2 Value at Risk

In practice, developed and widely used method for measuring and managing risks is the value index of risk, also called Value at Risk (VaR). This method can be used to calculate capital requirements, financial risks management, integration of risks into one value etc. VaR represents the risk value, which with given probability α will not be exceeded during certain time period N, see Hull [4]. Mathematically VaR can be expressed as one sided quantile of distribution of profits and losses for certain time of holding, and it is determined based on certain historical period. It is a function which consists of two parameters: time horizon (N) and the significance level (α %). VaR is, from theoretical point of view, relatively simple and understandable conception; however, practical determination can be a significant statistical problem. No standard calculation method exists for the determination of VaR. The differences among individual methods lie especially in the methods of simulating changes of risk factors and in methods transforming risk factors to change portfolio's value. Alexander [1] shows, that there are three basic methods used to calculate VaR in practice: **Variance and covariance method**, which is used for estimation of potential portfolio's losses volatility and correlation, which are acquired from historical data, **Historical simulation**, where the potential future loss is estimated based on losses which happened in the past, **Monte Carlo simulation**, which works with large number of simulations of portfolio's value development and which will be further described in detail.

3 Monte Carlo simulation

Monte Carlo simulation (MC) is a flexible tool modeling stochastic processes and is used to determine the value of non-linear instruments or can be used where mathematical methods fail, see Alexander [1]. Method is deriving from the law of big numbers; where the large numbers of randomly generated risk factors with selected characteristics come close to theoretical assumption, see Tichý [7]. When executing Monte Carlo simulation the following procedure can be used; based on selected probability distribution (e.g. Gaussian, Poisson, Student's distribution etc.), a vector of random numbers is generated, see below. In case that the portfolio does not contain more files, it is necessary to estimate also correlation structure by e.g. Cholesky algorithm. Subsequently development of yield (x) of assets with selected model specifying behavior of individual portfolio instruments is simulated e.g. Brownian motion, Levy's model etc. Specifically for Brownian motion, which will be used in the paper, the development of yields can be defined as Tichý [7]:

$$x^i = \mu \cdot \Delta t + \sigma \cdot \tilde{z}^i \cdot \sqrt{\Delta t}\,, \tag{1}$$

where μ is average yield , σ is standard deviation, $\tilde{z}$ is a random number from normalized normal distribution $N(0;1)$, Δt is increase of time, i expresses i-th asset.

For generation of random numbers can be selected various probability distributions. In this case will be used to normal and student probability distribution. Normal distribution is characterized by two parameters; mean value (μ) and standard deviation (σ). The next will be used to student distribution, which is used in finance as probabilistic models assets returns. This distribution is characterized by three parameters namely mean value (μ), standard deviation (σ) and degrees of freedom (υ). The student distribution becomes closer to the normal distribution when the parameter υ increase.

As it was mentioned above, MC is based on generating large numbers of random scenarios, whose selected characteristics will come close to theoretical assumption. Estimation's error than corresponds to standard deviation of result. In 1997 Boyle et al. [2], introduced techniques, which are trying to lower estimation error (dispersion of result) and by this increasing simulation's Monte Carlo effectiveness. At the same time, it comes to reducing the number of generated scenarios and decreasing the time requirement of Monte Carlo simulation. Among these procedures of MC belong: Antithetic Sampling Monte Carlo, Stratified Sampling Monte Carlo, Control variants Monte Carlo, Moment matching Monte Carlo and others.

When applying **Primitive Monte Carlo simulation** (PMC) random elements are generated so that they correspond to characteristics of selected probability distribution. This technique is relatively quick, but the estimation will be sufficiently accurate only for large number of random scenarios. Large numbers of scenarios lead to higher time demand of simulation see Tichý [7]. From this reason technique which enables to improve effectiveness of PMC, can be applied; e.g. Glasserman [3] or Jäckel [5] are dealing with this.

Antithetic Sampling Monte Carlo (ASMC) is for its simplicity and comprehensibility used very often in finance. The method is based on negative correlation among vectors of random elements, meaning $\rho(X, \overline{X}) = -1$, see Tichý [7]. Supposing X random elements $\tilde{z}$ from normalized normal distribution then by multiplying vector with coefficient *-1* we will get vector $\overline{X}$ of random elements $-\tilde{z}$. By unification of both vectors we can achieve double the amount of random elements, which better fulfill characteristics of selected distribution. It is a method, which leads to the decrease of the time consumption and achievement of zero mean, meaning the symmetry of probability distribution. The method's limitations lie in the fact that it can be used only when generating random elements from symmetric probability distributions.

Latin hypercube sampling (LHS) represents a procedure which efficiently extends stratified sampling to random vectors whose components are independent. The probability is divided into sub-intervals, and for each variable the sequence of sub-intervals are permuted separately. This method enables generating two or more mutually independent groups of random numbers and can be used for generating processes, which consist of random numbers from the distributions with various characteristics. It is generated *i* independent samples (*K1…Kd*) and is generated *m* independent permutations (*P1…Pd*) of {*1…n*} drawn from the distribution that makes all permutations equally probable. LHS is given by (Packham and Schmidt, [6]):

$$LHS = \frac{P_i^j - 1}{n} + \frac{K_i^j}{n} \quad j = 1,....d, \quad i = 1,...n \tag{2}$$

This method can be with dependence (LHSD). Essentially, LHSD extends LHS to random vectors with dependent components. Proceed as above practice, but correlation matrix must be equal correlation matrix of sort generated component.

4 Estimation of capital requirements

In the case of internationally oriented market subjects the role of currency risks plays an important role; this derives from unexpected changes of foreign currency exchange rates and as a result it changes the market price and also the subject market's position. Since insurers invest especially on European and American markets, the daily closing prices of three stock indexes, which are denominated in three different currencies, were used as input data: Down Jones Industrial Average (DJI) denominated in USD, Deutsche Aktien IndeX (DAX) denominated in EUR and FTSE 100 (FTSE) denominated in GBP. Prices of individual indexes were inquired on daily basis in the period of January 1st, 2003 to January 1st, 2014 (D.).[2] For the same period exchange rates of foreign currencies to CZK were also inquired. The exchange rates announced by Czech National Bank[3] were considered. In case that some stock markets did not trade, the missing data were substituted by the value from previous business day. We have available time line of 2851 daily logarithmic yields of stock markets and exchange rates.

Further, we suppose market portfolio (M), which was determined, based on conditions of Markowitz model and its composition is shown in Table 1.

	DJI (%)	**DAX (%)**	**FTSE (%)**	**E(R_D) (%)**	**σ(R_D) (%)**	**S**	**K**
M	0.3027	0.2699	0.4274	0.0138	1.0150	0.1315	9.9316

Table 1 Market portfolio´s composition

From the aforementioned Table 1 it is evident that the portfolio do not has normal distribution. The mean value is approximately 0.01 % and standard deviation is 1.02 % for market portfolio. The data are more or less symmetrically distributed around mean value; there is identified low positive skewness (*S*) and higher kurtosis (*K*) compared to normal distribution. Higher kurtosis is typical for financial assets. Between indexes there is slight dependence, the most correlation is between DJI/DAX about 0.5. This information is used to for calculation of Cholesky´s matrix.

VaR will be calculated by methods PMC, ASMC and LHSD, at the significance levels and time horizon, which correspond to solvency capital requirement (SCR) and minimal capital requirement (MCR) according to Solvency II, meaning at 0.5% and 15% significance level for a yearlong time horizon. We suppose that yields

[2]Data available at: <http://finance.yahoo.com/>

[3]Data available at :< http://www.cnb.cz/cs/index.html>

have multivariate normal distribution and multivariate student distribution with 4.5 degrees of freedom, behavior of individual portfolio's instrument follows Brownian motion and 100 000 random scenarios are generated. Subsequently, used methods will be compared. Then, the data of initial period (D) will be divided into four parts based on the development of economic situation on market, and capital requirements will be compared based on these periods. First period (I.) is historic series for period from 1. 1. 2003 to 31. 12. 2006 (before economic crises), next part (II.) is period from 1. 1. 2007 to 29. 8. 2008 (beginning of economic crises), third period (III.) includes data from 1. 9. 2008 to 31. 12. 2010 (economic crises) and last part (IV.) is period from 1. 1. 2011 to 1. 1. 2014 (retreat of economic crises). Calculations were performed in Wolfram Mathematica 9.0.

By PMC, ASMC and LHSD are estimated probability distributions for market portfolio and determined descriptive characteristics, which is shown in Table 2.

	PMC		**ASMC**		**LHSD**	
	$\mathbf{E(R_p)^*}$	$\mathbf{\sigma(R_p)^*}$	$\mathbf{E(R_p)^*}$	$\mathbf{\sigma(R_p)^*}$	$\mathbf{E(R_p)^*}$	$\mathbf{\sigma(R_p)^*}$
N (μ,σ)	0.0171	1.0166	0.0138	1.0094	0.0171	1.0166
S (μ,σ,υ)	0.0190	1.3657	0.0138	1.3962	0,0190	1.3656
	S	**K**	**S**	**K**	**S**	**K**
N (μ,σ)	0.0067	3.0103	-0.000	3.0028	0.0067	3.0103
S (μ,σ,υ)	0.0842	9.1261	-0,000	9.5763	0.0842	9.1261

Table 2 Descriptive characteristics for probability distribution

From the above table it can be seen that individual estimations based on multivariate normal distribution does not correspond with empirical values. It differs in all monitored characteristics, especially in kurtosis, which is very low. This estimation is very close to standard normal distribution, which is characterized by zero mean and standard deviation of one. Therefore student distribution is applied. Individual estimation based on multivariate student distribution more correspond with empirical values for all methods especially at mean value and kurtosis, but standard deviation increased by approximately 0.35 p.p. and also mildly increased mean value. The kurtosis can control by parameter υ (degree of freedom). In this case is used to $\upsilon = 4.5$, but should be selected lower value, because this change leads to increase kurtosis. For estimation it is better to use multivariate student distribution, which is able to capture higher kurtosis and heavy tails, which is characteristic for financial assets.

A comparison of different methods shows that we cannot state which method is the best. The best estimation of mean value is acquired by ASMC method for both probability distributions. At the same time the advantage of this method is its simplicity and low time consuming. Its limitations lie in the possibility to utilize it only for symmetric distribution of probability. More precise estimate of standard deviation is achieved by PMC and LHSD method which also better estimate the skewness. Better estimation of kurtosis for multivariate normal distribution is again provided by PMC and LHSD methods. ASMC method has better results for multivariate student distribution. PMC and LHSD method in our case lead to the same results from which we can assume that with sufficient number of scenarios it will provide similar results as other used more sophisticated approaches LHSD method is more time consuming in comparison with the other methods and in this case it is the least appropriate.

Based on selected data and aforementioned procedures the capital requirements for currency risk were calculated for various time periods. The results are shown in Table 3.

	Multivariate normal distribution						**Multivariate student distribution**					
	PMC		**ASMC**		**LHSD**		**PMC**		**ASMC**		**LHSD**	
	SCR	MCR	SCR	MCR	SCR	MCR	SCR	MCR	SCR	MCR	SCR	MCR
D.	37.21	12.35	37.66	13.07	37.21	12.35	62.81	14.10	66.65	15.91	62.81	14.10
I.	28.16	6.39	28.9	7.21	28.38	6.27	48.34	8.21	50.38	8.89	48.34	8.21
II.	53.10	30.57	53.28	31.23	39.78	14.38	73.35	32.15	75.06	33.03	73.35	32.15
III.	52.80	18.93	53.66	19.97	52.80	18.93	83.57	21.34	85.88	22.60	83.57	21.34
IV.	22.87	2.06	23.29	2.68	22.87	2.06	41.76	3.52	43.12	4.30	41.76	3.52

Table 3 Capital requirements for market portfolio and selected periods in %

It is very important to realize that with growing level of significance also grows the value of VaR and by this also capital requirements value. The same stands for time horizon, for which VaR is calculated, with growing time horizon the requirements for held capital are growing. As a result of this, SCR will always reach higher values then the MCR because it is computed at a higher significance level.

From Table 3 it is evident that for the whole period (D) and multivariate normal distribution the SCR is oscillating at approx. 37 % and MCR at approx. 12 % for all used methods. With the multivariate student distribution the values of SCR increased approx. about 25 pp and valued MCR approx. about 2 pp.

From the comparison of individual periods which present different stages of economic development (I.-IV.) it is evident that the capital requirements increase in the period of economic crises when the subject is exposed to higher risk and vice versa. From the multivariate normal distribution results we can see that in the period of stable economic situation I. the SCR is oscillating at 28 % and MCR at 6-7 % with all methods. In the following period II. the requirements for capital raised approximately about 25 pp In the period of economic crises III the value of SCR oscillate at a similar level as in a previous period with methods PMC and ASMC, but value of MCR decreased about 12 pp. which does not correspond to a theoretical assumptions. In the last period IV. the capital requirement dropped below the requirement level in the period before crises. Estimating the capital requirements by LHSD method more correspond to theoretical basis. There is not a problem with the decrease of capital requirements between the II. and III. period as it was with the previously mentioned two methods. In the period of stable economic situation I. the SCR reaches the 28 % and MCR 6 % level, when economic crises II. started the SCR increased about 12 pp and MCR about 8 pp In the third period the requirements on capital requirements grew again to 53 % and 19 %. In the period of economic crises decline IV. Both capital requirements decreased.

From comparing the requirements on held capital acquired by utilizing the multivariate student distribution derives that the results are very similar with all used methods. In the first period I. the SCR is oscillating approx. at 48 % and MCR at 8 %. In the period of economic crises III. the SCR grew to level 83 % and with ASMC method to even higher 85 %. MCR in this period reaches level 33 % which means a decrease in comparison with a previous period II. In the last period the values are coming back again and that is below the level of held capital in the period of economic situation I. Utilizing the multivariate student distribution leads to higher capital requirements in comparison with the utilization of normal distribution especially because it can record the skewness which is characteristic for financial assets.

5 Impacts of capital requirements on the insurance companies

The implementation of Solvency II will lead to many changes in the insurance companies economic activity. Insurance companies have been already undergoing some changes within the preparation for Solvency II implementation. This process is especially demanding for insurance mathematicians (models calculations and work with data), from the viewpoint of internal processes, requirements for information systems etc. According to the new methodology the capital requirements usually reach higher values than in the current regime. It also depends on the insurance company type and also on the country in which the insurance company is based. In the Czech environment the highest values of capital requirements are reached by life, non-life and market risks.

Capital requirements affect the items of insurance company's balance sheet, the fact if individual items will grow or decrease cannot be definitely stated, again it depends on the type of insurance company. It is very important to realize that everything is related and any movement on one side of the balance sheet will most likely be balanced by movement of another item. At the side of liabilities the increase of capital requirement will most likely lead to increase of equity capital, however it cannot be stated if the increase will happen with all insurers. This increase can cause increase of investment e.g. foreign investment which can lead to higher monetary risks and consequently SCR will increase again. At the side of assets the insurers can execute changes in their own investment policies. It is very possible that it will come to the increase of investment into fixed income assetsat the expense of investment into property and other risk assets so that the insurance companies reach lower SCR. In case that the insurance company will have it its portfolio high level of risk assets it will have to balance this risk by its own resources which can show in the insurance price increase (price differentiation in an insurance sector). The same effect can have the capital requirements for non liquid assets. It is also possible that in the insurance effort to lower SCR it will come to the limitation of investment into long-term infrastructure projects. Other affected item probably will be all hedging techniques. Better security leads to the decrease of SCR however at the same time the requirements for insurance company's resources increases because better security presents other costs.

Capital requirements will definitely affect cost of capital (CoC) of the given institution. CoC can be defined as a loss of profits from investments which are not realized from the capital that the insurance company must keep in order to fulfill its liabilities. The CoC is acquired by multiplying average costs of equity capital in given business sector and capital requirements. CoC will then progress in the same direction as capital requirements because they are determined as a percentage from capital requirement, i.e. if the insurer holds higher amount of capital then the CoC will be higher and vice versa. Next affected index is the solvency ratio and SCR plays a

significant role in this index. In case that there will be no changes of own resources and SCR will grow then this index would decrease.

6 Conclusion

Because of the pressure on the transparency, trustworthiness and stability of financial system there is a great emphasis on correct and adequate risk management of individual subjects. Insurance sector is currently preparing for the implementation of new solvency regime Solvency II, which should insure certain guarantee of risk of given financial subject. Important area which the solvency framework is amending are the capital requirements, meaning requirements for holding certain level of capital so that the subject is able to permanently meet its requirements and liabilities.

Main aim of this paper was to determine capital requirements in the insurance sector by various methods of Monte Carlo simulations. Capital requirements were determined as a Value at Risk at the significance level 0.5% for SCR and 15% for MCR. Monte Carlo simulation, Antithetic Sampling Monte Carlo and Latin Hypercube Sampling Monte Carlo with dependence has been used.

It cannot be positively stated which selected method provides better estimate of probability distribution, because every one of them achieves better results with different parameters of probability distribution. However we can state that if there are sufficient numbers of scenarios the PMC achieves similar results as much more sophisticated methods and at the same time it is less time consuming. LHSD is a sophisticated method which is more time consuming and in this concrete example it provides the same estimate as PMC method, so its utilization is the least appropriate.

The implementation of new regime will influence the economic activity of insurance companies and their approach to risk. From up to now executed quantitative researches it is evident that the capital requirements according to Solvency II will be higher than in the current regime. Increased capital requirement will certainly affect the items of insurance company's balance sheet. The insurance companies can balance increased requirements on capital by the increase of their own assets which will reflect in the cost of capital and also in the price of the insurance. The increase of capital requirements can lead the insurance companies to make to effort to lower them again. The decrease of SCR can be executed by e.g. the change in investment strategy (moving from risk investment into less risky ones or substituting illiquid investment by more liquid assets, higher diversification), or by implementing better security, which will again cause additional costs etc.

Acknowledgements

This work was supported by the SGS project of VSB-TU Ostrava under No. SP2014/16 and Czech Science Foundation through project No. 13-13142S. The support is greatly acknowledged.

References

[1] Alexander, C.:*Market Risk Analysis Volume IV: Value at Risk Models*. John Wiley & Sons Inc.,Chichester, 2008.
[2] Boyle, P., Broadie, M., Glasserman, P.: Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control* **21** (1997), 1267-1321.
[3] Glasserman, P.: *Monte Carlo Methods in Financial Engineering*. Springer, Berlin, 2004
[4] Hull, J. C.: *Risk Management and Financial Institutions*. Pearson Prentice Hall, Upper Saddle River (USA), 2007.
[5] Jäckel, P.: *Monte Carlo methods in finance*. John Wiley & Sons Inc., Chichester, 2002.
[6] Parkham, N., Schmidt, W.: *Latin hypercube sampling with dependence and applications in finance.* Frankfurt: Frankfurt School of finance & Management, 2008.
[7] Tichý, T.: *Simulace Monte Carlo ve financích: Aplikace při ocenění jednoduchých opcí*. VŠB-TU Ostrava, Ostrava, 2010.
[8] European Insurance And Occupational Pensions Authority.: EIOPA [online]. 2010 [cit. 2014-01-21]. Available at: < https://eiopa.europa.eu/>.

An Alternative Approach to Interval AHP

Jiří Mazurek[1]

Abstract: Since 1980s the analytic hierarchy process (AHP) became the standard tool for a group and multiple criteria decision making. The core of AHP consists of pair-wise comparisons of alternatives and criteria with regard to a superior element in a given hierarchy. When an uncertainty is present in a decision making interval or fuzzy AHP is usually considered. The aim of this paper is to propose a new method for interval AHP where a relative importance of one element (alternative or a criterion) over another element is expressed by an interval rather than a crisp number from Saaty's fundamental scale from 1 to 9. In the proposed method weights of alternatives (criteria) are estimated by an interval generalization of the geometric mean, and for the final ranking of alternatives a formula based on a possibility theory is employed. The proposed method is computationally simpler than the standard linear programming (LP) approaches and its use is demonstrated on examples. Also, results of the proposed method are compared to those obtained by LP.

Keywords: AHP, geometric mean, interval AHP, multiple criteria, uncertainty.

JEL Classification: C44, D80.
AMS Classification: 90-08.

1. Introduction

Since 1980s the analytic hierarchy process (AHP) by T. L. Saaty became the standard tool for a group and multiple criteria decision making, see [6] or [7]. The core of AHP consists of pair-wise comparisons of alternatives and criteria with regard to a superior element in a given hierarchy with the use of Saaty's fundamental scale $\{1/9,1/8,...,1,...,8,9\}$.

When an uncertainty is present in a decision making interval or fuzzy AHP is usually considered. In this paper the first approach will be followed. Interval AHP was proposed by Saaty and Vargas in [8], they used Monte Carlo simulations for derivation of weights. Later other methods based on linear programming were proposed, such as Lexicographic Goal Programming (LGPAHP) in [4], multiplicative AHP (MAHP) in [1], [5] or [12], or complete enumeration and upper approximation in [2], [3] and [9].

The aim of this paper is to propose a new method for interval AHP where a relative importance of one element (alternative) over another element is expressed by an interval rather than a crisp number from Saaty's fundamental scale $\{1/9,1/8,...,1,...,8,9\}$. In the proposed method weights of alternatives (criteria) are estimated by an interval generalization of the geometric mean, and for the final ranking of alternatives a formula based on the possibility theory is employed. The proposed method is computationally simpler than the standard linear programming (LP) approach and its use is demonstrated on examples. Also, results of the proposed method are compared to those obtained by LP.

The paper is organized as follows: section 2 provides preliminaries, in section 3 the proposed method is described, and in section 4 illustrative examples are provided. Conclusions close the article.

2. Preliminaries

In this paper interval AHP problem with three levels: a goal, criteria and alternatives will be considered. Let *n* be the number of alternatives and let *k* be the number of criteria. A decision maker expresses a relative importance of criteria and alternatives as real intervals (in this paper it is assumed that these intervals conform to the Saaty's fundamental scale) which form the following interval comparison matrices:

[1] School of Business Administration in Karviná, , Silesian University in Opava, Department of Mathematical Methods in Economics, Univerzitní náměstí 1934/3, Czech Republic, mazurek@opf.slu.cz

$$\begin{bmatrix} 1 & [l_{12},u_{12}] & [l_{13},u_{13}] & \dots \\ \dfrac{1}{[l_{21},u_{21}]} & 1 & [l_{23},u_{23}] & \dots \\ \dfrac{1}{[l_{31},u_{31}]} & \dots & 1 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (1)$$

If weights of alternatives or criteria are assumed to be crisp numbers $w_i \geq 0, \forall i$, then they satisfy relation $\sum_i w_i = 1$. If weights of alternatives (or criteria) $w_i = \left[\underline{w_i}, \overline{w_i}\right]$ are assumed to be intervals, then it should satisfy relations for interval probabilities, see e.g. [3] or [11]:

$$\sum_{i \neq j} \overline{w_i} + \underline{w_j} \geq 1; \forall j$$

$$\sum_{i \neq j} \overline{w_j} + \underline{w_i} \leq 1; \forall j$$

These conditions are natural constraints for linear programming approaches to interval AHP.

For example Lexicographic Goal Programming (LGPAHP) starts with an interval comparison matrix (as in Figure 1), where interval bounds satisfy relations: $l_{ij} \leq u_{ij}, u_{ij} = 1/l_{ji}, \forall i, j$, unknown weights w_i lie within appropriate intervals: $l_{ij} \leq \dfrac{w_i}{w_j} \leq u_{ij}$, and these inequalities can be transformed into the following inequalities: $l_{ij} w_j - p_{ij} \leq w_i \leq u_{ij} w_j + q_{ij}$, $i = 1,2,...,n-1$, $j = i+1, i+2,...,n$ [12]. Values of p_{ij} and q_{ij} are non-negative and equal to 0 if all judgments are consistent. Finally, the following function must be minimized:

$Z = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(p_{ij} + q_{ij} \right)$ [12].

In other LP approaches the function $\sum_i \left(\overline{w_i} - \underline{w_i} \right)$ (the sum of interval widths) or more sophisticated function under appropriate constraints is minimized.

2.1. Interval arithmetics and interval generalization of the geometric mean

Arithmetics of real numbers can be easily extended to arithmetics on real intervals. Let $a,b,c,d \in R$ and $I_1 = [a,b]$ and $I_2 = [c,d]$ be intervals. Then:

$$I_1 \pm I_2 = [a \pm c, b \pm d] \quad (2a)$$

$$I_1 \cdot I_2 = [a \cdot c, b \cdot d] \quad (2b)$$

$$\frac{I_1}{I_2} = \frac{[a,b]}{[c,d]} = \left[\frac{a}{d}, \frac{b}{c}\right], \; c,d \neq 0 \quad (2c)$$

With the use of these relations (2a-c) the geometric mean for intervals can be defined. Let I_i be real intervals, then interval geometric mean is defined as follows:

$$\bar{I}_G = \sqrt[n]{\prod_{i=1}^{n} I_i} \quad (3)$$

2.2 Aggregation of weights

In the proposed method it is assumed that weights of criteria and alternatives are computed from interval comparison matrices via relation (3) and have the form of real intervals.

It should be noted that weights obtained from interval geometric mean do not satisfy interval probability conditions (they are not normalized to 1), in this sense they are rather 'pseudoweights', as they are just values needed for a dominance relation (5),

Let v_i be weights of criteria and let w_i^k be a weight of an alternative *i* with regard to a criterion *k*. Then the weight of an alternative *i* with regard to a goal is given as:

$$w_i = \sum_{k=1}^{K} w_i^k \cdot v_k \tag{4}$$

This value (pseudoweight) enters the dominance relation (5), so alternatives can be ranked from the best to the worst (but generally only quasi-ordering of alternatives is provided as some alternatives might be tied).

2.3 Comparing of interval numbers

Since 1980s many procedures for ordering of interval numbers or fuzzy numbers were proposed. In this paper a formula based on the possibility theory proposed by Xu in [10] originally in the context of comparing uncertain linguistic variables is used for ranking interval numbers.

Definition 1. ([10]). Let $I_1 = [a,b]$ and $I_2 = [c,d]$ be intervals of real numbers. Then the degree of possibility that $I_1 \geq I_2$ is defined as:

$$p(I_1 \geq I_2) = \max\left\{1 - \max\left(\frac{d-a}{b+d-a-c}, 0\right), 0\right\} \tag{5}$$

Relation (5) has some desirable properties [10]:

i) $0 \leq p(I_k \geq I_l) \leq 1, \forall k,l$,

ii) $p(I_k \geq I_l) + p(I_l \geq I_k) = 1$, $\forall k,l$,

iii) If $p(I_k \geq I_l) \geq 0.5$ and $p(I_l \geq I_m) \geq 0.5$ then $p(I_k \geq I_m) \geq 0.5$.

Therefore, relation (5) is nothing else than a dominance relation on intervals which can be used for a comparison of alternatives in interval AHP, where compared intervals are weights of alternatives. Trivially, if $I_1 = [a,b]$, $I_2 = [c,d]$ and $c > b$ (intervals are non-overlapping), then $I_2 > I_1$.

3. The proposed method

The proposed method for interval AHP uses interval arithmetics, interval generalization of the geometric mean and a formula (5). It proceeds in the following six steps:

1. Importance of all criteria with regard to a goal is expressed by an interval matrix *C*.
2. All alternatives are compared with regard to all criteria by interval matrices A_k.
3. Interval weights of criteria from the matrix *C* are obtained with the use of relation (2).
4. Interval weights of all alternatives with respect to all criteria are obtained with the use of relation (2).
5. Aggregated weights (with regard to a goal) of all alternatives are obtained by relation (4)
6. In this final step alternatives' aggregated weights are compared via relation (5) and (a weak) ordering (or ranking) of all alternatives with regard to a goal is achieved.

The proposed method is natural and computationally simple; it does not require special computer software. On the other hand the method does not provide weights of alternatives, but only possibilities that a given alternative is (not) better than some other alternative.

Also, in the final step all alternatives should be compared pair-wise, which might be time expensive for larger number of alternatives. Furthermore, an evaluation of consistency of judgments for interval matrices is still an open question too.

4. Numerical examples

Example 1. The following interval comparison matrix of four alternatives is from [4] and [9]:

$$\begin{bmatrix} 1 & [1,2] & [1,2] & [2,3] \\ [1/2,1] & 1 & [3,5] & [4,5] \\ [1/2,1] & [1/5,1/3] & 1 & [6,8] \\ [1/3,1/2] & [1/5,1/4] & [1/8,1/6] & 1 \end{bmatrix}$$

Solution:
Accorrding to LGPAHP method the weights are as follows. For the upper triangular matrix we obtain weights $w=(0.303,0.455,0.151,0.091)$, so the ranking of weights is $w_2 \succ w_1 \succ w_3 \succ w_4$, but for the lower triangular matrix different weights are obtained $w=(0.364,0.364,0.182,0.090)$, so now the ranking of weights is $w_2 = w_1 \succ w_3 \succ w_4$ (notice the large change in the weight of w_2). For the whole matrix one obtains the ranking $w_2 = w_1 \succ w_3 \succ w_4$ as well.
By the MAHP method by Yu et al. in [12], only one vector of weights (and one ranking) is attained, in this case $w=(0.281,0.469,0.156,0.094)$, so the ranking of weights is $w_2 \succ w_1 \succ w_3 \succ w_4$.
By the proposed method (see section 3) we get interval weights for all four alternatives: $w=([1.189,1.860],[1.565,2.236],[0.880,1.278],[0.302,0.380])$. With the use of formula (5) we compare all alternatives (or their weights – intervals respectively):
$p(w_1 \geq w_2)=\max\left\{1-\max\left(\frac{2.236-1.189}{1.860+2.236-1.189-1.565},0\right),0\right\}=0.220$, so $w_2 \succ w_1$, etc. At the end we obtain the final ranking $w_2 \succ w_1 \succ w_3 \succ w_4$, which is the same as in MAHP.

Example 2. Consider two criteria C_1 and C_2, and four alternatives *A*, *B*, *C* and *D*. A comparison of relative importance of criteria and alternatives is provided in Tables 1, 2 and 3. The goal is to find the best alternative.

Solution:
On the left hand side of Tables 1-3 geometric means (2) are provided. For each alternative agreggated weights via (4) are estimated:

$$w_A=\sum_{k=1}^{2} w_A^k \cdot \upsilon_k = [1.414.,1.732]\cdot[1.414,1.732]+[0.577,0.707]\cdot[1.968,2.783]=[3.186,4.968].$$

$$w_B=\sum_{k=1}^{2} w_B^k \cdot \upsilon_k = [1.414,1.732]\cdot[0.508,0.639]+[0.577,0.707]\cdot[1.107,1.681]=[1.357,2.295].$$

$$w_C=\sum_{k=1}^{2} w_C^k \cdot \upsilon_k = [1.414.,1.732]\cdot[3.201,3.722]+[0.577,0.707]\cdot[0.562,0.904]=[4.852,7.086].$$

$$w_D=\sum_{k=1}^{2} w_D^k \cdot \upsilon_k = [1.414.,1.732]\cdot[0.354,0.467]+[0.577,0.707]\cdot[0.380,0.508]=[0.720,1.168].$$

In the final step all alternatives are compared pair-wise. Trivially we obtain: $A \succ B \succ D$ and $C \succ B \succ D$ (because interval weights of these alternatives are non-overlapping intervals). Only pair A-C is endowed by non-overlapping inervals, so with the use of the relation (5) we obtain the possibility that *A* is better (dominates) *C* as follows:

$p(A \geq C)=\max\left\{1-\max\left(\frac{7.086-3.186}{4.968+7.086-3.186-4.852},0\right),0\right\}=0.029$, hence $p(C \geq A)=0.971$ and $C \succ A$

Finally, the ranking of al alternatives is obtained: $C \succ A \succ B \succ D$. An alternative *C* is the best.

$$\begin{bmatrix} 1 & [2,3] \\ [1/3,1/2] & 1 \end{bmatrix} \begin{bmatrix} 1.414,1.732 \\ 0.577,0.707 \end{bmatrix}$$

Table 1 A comparison of both criteria C_1 and C_2

$$\begin{bmatrix} 1 & [2,3] & [1/3,1/4] & [3,4] \\ [1/3,1/2] & 1 & [1/5,1/6] & [1,2] \\ [3,4] & [5,6] & 1 & [7,8] \\ [1/4,1/3] & [1/2,1] & [1/8,1/7] & 1 \end{bmatrix} \begin{bmatrix} 1.414,1.732 \\ 0.508,0.639 \\ 3.201,3.722 \\ 0.354,0.467 \end{bmatrix}$$

Table 2 A comparison of all alternatives with regard to the criterion C_1

$$\begin{bmatrix} 1 & [1,2] & [3,5] & [5,6] \\ [1/2,1] & 1 & [1,2] & [3,4] \\ [1/5,1/3] & [1/2,1] & 1 & [1,2] \\ [1/6,1/5] & [1/4,1/3] & [1/2,1] & 1 \end{bmatrix} \begin{bmatrix} 1.968,2.783 \\ 1.107,1.681 \\ 0.562,0.904 \\ 0.380,0.508 \end{bmatrix}$$

Table 3 A comparison of all alternatives with regard to the criterion C_2

Conclusions

The aim of the paper was to propose an alternative to linear programming (LP) approach to the interval AHP, which is based on the use of interval arithmetic and interval generalization of the geometric mean. The proposed method is straightforward and doesn't require any mathematical software so it might be a suitable method in practice. Alternatively, it can be used as a complement to the LP methods as problems of multiple-criteria decision making are usually solved by more than one method (because there is no ideal one). Results of the proposed method are likely to be similar to results obtained by the LP methods but the degree of similarity among these methods in general has to be examined yet. Also, a consistency of preferences in this framework has to be investigated; hence further research might focus on the problems stated above.

Acknowledgement

The paper was supported by the grant of the Czech Grant Agency no. 14-02424S.

References

[1] Chandran, B., Golden, B, Wasil, E.: Linear programming models for estimating weights in the analytic hierarchy process, *Computers & OperationsResearch* **32**(9) (2005), 2235-2254.

[2] Cox, M. A. A.: Examining alternatives in the interval analytic hierarchy process using complete enumeration, *European Journal of Operational Research* **180**(2) (2007), 957-962

[3] Entani, T., Tanaka, H.: Interval Estimations of Global Weights in AHP by Upper Approximation, *Fuzzy Sets and Systems* **158**(17), 1913-1921.

[4] Islam, R., Biswal, M. P., Alam, S. S.: Preference programming and inconsistent interval judgments, *European Journal of Operational Research* **97**(1) (1997), 53-62.

[5] Ramanathan, R.: Stochastic decision making using multiplicative AHP," *European Journal of Operational Research* **97**(3) (1997), 543-549.

[6] Saaty, T. L.: *The Analytic Hierarchy Process*. McGraw Hill, New York, 1980.

[7] Saaty, T. L.: Decision making with the analytic hierarchy process. *Int. J. Services Sciences* **1** (2008), 83–98.

[8] Saaty, T. L., Vargas, L. G.: Uncertainty and rank order in the analytic hierarchy process, *European Journal of Operational Research* **32**(1) (1987), 107-117.

[9] Wang, Y. M. , Yang, J. B., Xu, D. L.: Interval weight generation approaches based on consistency test and interval comparison matrices, *AppliedMathematics and Computation*, **167**(1) (2005), 252-273.

[10] Xu, Z.: A Direct Approach to Group Decision Making with Uncertain Additive Linguistic Preference Relations, Fuzzy Optimization and Decision Making **5**(1) (2006), 21-32.

[11] Yager, R. and Kreinovich, V.: Decision Making under Interval Probabilities, *Departmental Technical Reports (CS).* Paper 433, 1999.

[12] Jing Rung Yu, Yu-Wen Hsiao, and Her-Jiun Sheu: A Multiplicative Approach to Derive Weights in the Interval Analytic Hierarchy Process, *International Journal of Fuzzy Systems*, **13**(3) (2011), 225-231.

On a Gravity Equation of Trade: A Case of Germany and the Czech Republic

Jiří Mazurek [1], Lucie Zotyková [2]

Abstract: Since 1960s gravity models or equations (economic analogies of Newton's Law of gravity) of various kinds are used in modeling volumes of international trade with a significant statistical success. These models assume that a volume of a bilateral trade is positively proportional to 'sizes' of trading economies and negatively proportional to their distance. The aim of the paper is to evaluate exports from Germany (the 3rd world's exporter) and the Czech Republic (small and open post-transition economy used for a comparison) to their international partners via frictionless aggregate gravity equation with the use of the latest (2012) export data. Results of linear regression with corrected heteroscedasticity (for both countries separately) indicate that the model predicts exports successfully with adjusted determinacy coefficients around 0.75 (GER) and 0.98 (CZE) respectively. Also, coefficients associated with the gross domestic product (0.60 and 0.65) and distance between countries (1.24 and 0.78) were found close to values of other similar empirical studies. An extension of the model including trade frictions associated with the existence of national borders is provided as well.

Keywords: Czech Republic, export, Germany, gravity model, international trade.

JEL Classification: C51, F14, F17.
AMS Classification: 62J05

1 Introduction

The gravity model (equation) of international trade was used for the first time by Tinbergen in [13], and then it was followed by many other studies, see e.g. theoretical derivation of gravity equation in [2], and [5], [3], [10] or [4] to name a few. A concise review of gravity approach to international trade can be found e.g. in [1]. Gravity models theoreticaly explain the role of an economic size in bilateral trade flows at any scale (countries, regions, etc.), though the role of a distance is not well understood yet (see [7]), and it also applies to migration or direct foreign investments [1]. Influence of trade barriers such as borders among countries or different language or currency can be explained by gravity models as well.

The standard gravity model (equation) of international trade takes usually the following form [7]:

$$T_{ij} = k \frac{GDP_i^{\alpha} GDP_j^{\beta}}{d_{ij}^{\gamma}} \tag{1}$$

where T_{ij} is a trade from a country i to a country j, GDP_i denotes gross domestic product, k is a positive coefficient and d_{ij} is a geographic distance of both countries.

Anderson in [1] assumes that supply Y_i of a country i is attracted by a demand E_j of a country j, where d_{ij} denotes a distance of both countries, and proposes the following frictionless and aggregate gravity model of trade:

$$T_{ij} = \frac{Y_i E_J}{d_{ij}^{2}} \tag{2}$$

Also, in some alternative gravity models an income per capita (along with a population) of countries is used instead of supply and demand, and an error term is added on the right hand side of equations; see e.g. [2]. From

[1] Silesian University in Opava/School of Business Administration in Karviná, Department of Mathematical Methods in Economics, Univerzitní náměstí 1934/3, 733 40 Karviná, mazurek@opf.slu.cz.
[2] Silesian University in Opava/School of Business Administration in Karviná, Department of Mathematical Methods in Economics, Univerzitní náměstí 1934/3, 733 40 Karviná, zotykova@opf.slu.cz.

simple models (1) and (2) more sophisticated models with trade frictions (trade barriers) and disaggregated goods (frictions are different for different goods in reality) can be formulated, see [1].

The empirical evidence for gravity models is rather strong. A metaanalysis of 1467 estimates in 103 papers by Disdier and Head in [9] found that coefficients α, β and γ in (1) are close to 1. Remarkably, coefficient γ is stable (around 1) for more than one century of data. A theoretical explanation of this result can be found e.g. in [7].

In this paper the following gravity equation (model) for export shares of a given country will be considered:

$$E_{ij}(\%) = k_i \frac{GDP_j^{\alpha}}{DIST_{ij}^{\beta}} \tag{3}$$

In (3) E_{ij} denotes a share of an export (in %) from a country *i* to a country *j*, GDP_j is a gross domestic product of an importing country *j* (in billions of USD), $DIST_{ij}$ is a distance between countries *i* and *j* (given as an air distance of capital cities in kilometers), and α, β, and k_i are (positive) coefficients.

It should be noted that if absolute values of exports (e.g. in billions of USD) were considered in (3) instead of relative exports, then only coefficients k_i would change, which in turn wouldn't influence properties of regression models based on (3).

Relation (3) states that an export rises when an importing country is closer and/or richer. Because E_{ij} is given in (%), an export to all (*n*) trading partner countries must sum up to 100%. Also, from (3) it follows that two importing countries with equal GDP and distance from an exporting country *i* have the same share of export from this country. If an export was independent on a distance then export shares from a country *I* would be fully determined (proportional) by GDP of importing countries.

The aim of the article is to examine how the gravity model of exports (3) fits the export data for two selected Central European countries: Germany and the Czech Republic. These countries are suitably located in the middle of the continent surrounded by many trading partners in different distances unlike rather isolated countries such as Iceland or Cyprus. Furthermore, Germany is the third largest world exporter (just behind the USA and China), and it ranks among the most developed countries of the World. The Czech Republic's economy is rather an opposite: it is a small and open post-transition economy of the former Soviet bloc (moreover geographically close to Germany), so it can provide an interesting comparison with Germany. And for the last, but not least, there are available accurate and reliable data for exporting partners of both countries.

2 The data and the method

For the empirical investigation of the model (3) the following data were used:

- Exporting partners (shares in %) were obtained from [6], actualized in June 2013. Because the list of all exporting partners may be very long (including theoretically all countries of the World) and the data on partner countries with shares of only fractions of a percent are not sufficiently reliable and precise, the list was terminated when the sum of export shares of countries on the list exceeded 90%. The data on exporting partners for Germany and the Czech Republic is provided in Tables 1 and 2.
- Distances between a given country and its export partners (in km) were obtained from a distance calculator at [12]. The distance between two countries was defined as an air distance between their capital cities.
- GDP (PPP) in billion USD were retrieved from the Intl. Monetary Fund [11].

All the data for both countries are provided in Tables 1 and 2. It should be noted that export shares and GDP (PPP) change in time and are a subject of later revisions.

Exporter: CZECH REP.	EXPORT to (%)	DIST (km)	GDP (billion USD) PPP
Germany	32.390	280	3167
Slovakia	8.860	292	132
Poland	6.650	514	802
France	5.750	885	2252
United Kingdom	5.300	1034	2312
Austria	4.780	250	359
Russia	2.830	1664	2486
Italy	4.890	922	1813
Netherlands	3.940	707	710
Belgium	2.750	721	421
United States	1.760	6589	15653
Hungary	2.840	444	197
Spain	2.370	1773	1407
Switzerland	1.450	623	362
Sweden	1.820	1056	396
Romania	1.410	1076	274.1
Ukraine	0.970	1146	335

Table 1 The data for the Czech Republic as an exporter. Source: [6],[11], [12]

Exporter: GERMANY	EXPORT to (%)	DIST (km)	GDP (billion USD) PPP
France	9.9	879	2252
UK	6.91	932	2312
Netherlands	6.51	577	710
USA	7.96	6402	15653
Austria	5.4	523	359
Italy	6.59	1183	1813
China	3.69	7377	12261
Switzerland	4.15	752	362
Belgium	5.24	651	421
Poland	4.08	520	802
Spain	4.53	1870	1407
Russia	3.52	1616	2486
Czech Republic	2.85	280	287
Sweden	2.11	813	396
Hungary	1.82	691	197
Denmark	1.63	356	210
Turkey	1.62	2042	1125
Japan	1.36	8940	4575
Finland	0.96	1109	198
Korea	0.93	8150	1622
Slovakia	0.9	554	132
Brazil	0.9	9429	2230
Romania	0.87	1297	274.1

Exporter: GERMANY	EXPORT to (%)	DIST (km)	GDP (billion USD) PPP
India	0.86	5793	4716
UAE	0.83	4641	271
Portugal	0.82	2315	245
Norway	0.82	840	278
Greece	0.79	1804	281
South Africa	0.75	8831	579
Mexico	0.75	9741	1758

Table 2 The data for Germany as an exporter. Source: [6],[11], [12]

For the regression analysis relation (3) is reformulated in the following way:

$$E_{ij} = k_i \cdot GDP_j^{\alpha} \cdot DIST_{ij}^{\beta} \tag{4}$$

Taking natural logarithms (all variables are positive) of both sides of (4) yields:

$$\ln E_{ij} = \ln k_i + \alpha \cdot \ln GDP_j + \beta \ln DIST_{ij} \tag{5}$$

This model (with corrected heteroscedasticity) was analyzed with the use of the data (for each country separately) from Tables 1 and 2. For the regression statistical software Gretl was utilized.

3 Results

Results of linear regressions (5) for the Czech Republic and Germany (as exporters) are shown in Table 3. The model successfully models shares of export for both countries, as in the case of the Czech Republic both explanatory (independent) variables are highly significant ($p \leq 10^{-5}$) and coefficient of determination is around 0.98. Also in the case of Germany both explanatory variables are highly significant ($p \leq 10^{-3}$), but the coefficient of determination is smaller: 0.75. Generally, in both cases the model predicts exports shares successfully.

By the linear regression the following relationships were obtained:

- Czech Republic: $E_j = e^{5.177} GDP_j^{0.652} DIST_j^{-1.241}$ (6)
- Germany: $E_j = e^{2.544} GDP_j^{0.600} DIST_j^{-0.784}$ (7)

In (6) and (7) the subscript *j* denotes all importing countries (from Tables 1 and 2) with regard to exporters Germany and Czech Republic respectively. Values of coefficients α, β and k of the model (5) were found positive as expected for both countries. In both cases the coefficient α is close to 0.6, but β is higher in the absolute value than 1 in the case of the Czech Republic and smaller than 1 in the case of Germany. This result implies that a distance plays larger role in an export from the Czech Republic, or alternatively, that exporters from Germany are more capable of overcoming transportation distances (costs).

CZECH REP.	coefficient	GERMANY	coefficient
no. of observations	17	no. of observations	30
const.	5.17747***	const.	2.54425***
Log(DIST)	-1.2412***	Log(DIST)	-0.784429***
Log (GDP)	0.652561***	Log (GDP)	0.600304***

Table 3 Linear regression results for Germany and the Czech Republic

4 A gravity model with borders

Frictions of any kind (customs, borders, different languages or currency, bureaucratic obstacles, etc.) which influence real trade can be incorporated into gravity models, see e.g. [1].

The existence of national borders is considered one of the most important frictions in international trade. To account for possible trade frictions the model (3) is usually extended into the following form:

$$E_{ij}(\%) = k_i GDP_j^{\alpha} DIST_{ij}^{\beta} b_{ij}^{\gamma} \tag{8}$$

where b_{ij}^{γ} is a (minimal) number of borders between countries *i* and *j* and γ is a parameter. This model was examined with the use of the (same) data for Germany and the Czech Republic (for countries divided by a see the coefficient *b* was set to be 2 as a land-sea division is a natural border).

Results of linear regression are shown in Table 4. The coefficient γ was close to zero and not statistically significant in the case of the Czech Republic, but it was statistically significant (at $p = 0.05$) in the case of Germany. Values of coefficients in the model (8) were found to be similar to those for the model (4) without borders, see Table 3. Adjusted coefficient of determination for the Czech Republic: $R^2 = 0.971$, and for Germany: $R^2 = 0.859$. This value is slightly higher in the case of Germany, but almost the same for the Czech Republic.

Hence, the model with borders is not considerably better according to this (limited) study, so further research (both theoretical and empirical) on borders' influence on trade is needed. For example, borders within and outside of the Schengen area can be considered.

CZECH REP.	coefficient	GERMANY	coefficient
no. of observations	17	no. of observations	30
const.	5.381***	const.	2.319***
Log(DIST)	–1.332***	Log(DIST)	–0.813***
Log (GDP)	0.686***	Log (GDP)	0.687***
Log (b)	0.181	Log (b)	–0.289**

Table 4 Linear regression results for Germany and the Czech Republic: the model with borders

5 Conclusions

The aim of this paper was to examine exports from Germany and the Czech Republic to their trading partners by a frictionless aggregate gravity equation. Both countries were examined separately as only partners of a given country were considered for the evaluation. This is a slightly different approach from a standard methodology where volumes of trade among a group of countries are used as an input for gravity models.

Results of this study revealed that the frictionless gravity model (3) is very successful in fitting the data with adjusted determinacy coefficients equal to 0.75 for Germany and 0.98 for the Czech Republic respectively. Also coefficients associated with a GDP and a distance were found well within range of values determined by other similar studies. Also, it was found that the gravity model (8) with borders was only slightly better than the frictionless model in the case of Germany, but not better in the case of Czech Republic.

Further research might focus on other countries (especially outside Europe) or on trade frictions such as different language or currency.

Acknowledgements

One of authors, Lucie Zotyková, was supported by the project "Innovation of Educational Programs at Silesian University, School of Business Administration in Karviná" no. CZ.1.07/2.2.00/28.0017.

References

[1] Anderson J. E. The Gravity model. *NBER Working Paper Series*, pages: 47, 2010.

[2] Anderson J. E. A theoretical foundation for the gravity equation, *American Economic Review* 69 (1979), 106-116.

[3] Anderson, J. E., van Wincoop, E. Gravity with Gravitas, *American Econ. Review*, 93 (2003), 170-92.
[4] Baier, S. L., Bergstrand, J. H. Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Efects using the Gravity Equation, *Journal of Intl. Economics*, **77**(1) (2009), 77-85.
[5] Bergstrand J. H. The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence, *Review of Economics and Statistics*, **67**(3) (1985), 474-481.
[6] *Bridgat*. 2013. [online]. Available from: http://countries.bridgat.com.
[7] Chaney, T. The Gravity Equation in International Trade: an Explanation, *NBER Working Paper* No. 19285, 2013.
[8] Deardoff, A. V. Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World? In: *The Regionalization of the World Economy*, Chicago University Press, 1998.
[9] Disdier, A. C., Head, K. The Puzzling Persistence of the Distance Effect on Bilateral Trade, *The Review of Economics and Statistics*, MIT Press, **90**(1) (2008), 37-48.
[10] Helpman, E., Melitz, M.J. and Rubinstein, Y. Estimating Trade Flows: Trading Partners and Trading Volumes, *Quarterly Journal of Economics*, 123 (2008), 441-487.
[11] *International Monetary Fund* (*IMF*). (2013). [online]. Available from: http://www.imf.org.
[12] *Timeanddate*. 2013. [online]. Available from: http://www.timeanddate.com/worldclock/
[13] Tinbergen, J. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: The Twentieth Century Fund, 1962.

A Variable Neighborhood Search for the Capacitated Arc Routing Problem with Time Windows

Jan Melechovský[1]

Abstract. This paper presents a Variable Neighborhood Search (VNS) solution approach for the Capacitated Arc Routing Problem with Time Windows (CARPTW). The problem has received less attention compared to its vehicle routing counterpart. The problem is defined on a set of arcs or edges, a set of vehicles and a cost function mapping the arcs to real numbers. Each arc is characterized with a nonnegative demand, a nonnegative service time and a time window in which the service can start. The vehicles are assumed to be homogeneous with equal capacity. The solution to the problem consists of determining a minimum cost set of vehicle routes starting and ending at a depot node, such that the total demand of arcs is fully satisfied while the arcs time windows and vehicle capacity being respected. We propose a VNS metaheuristic to address the problem. The efficiency of our algorithm is evaluated on benchmarks from the literature.

Keywords: optimization, arc routing problem, metaheuristic, variable neighborhood search, time windows.

JEL classification: C61, D85
AMS classification: 05C51

1 Introduction

The Capacitated Arc Routing Problem (CARP) is an optimization problem of servicing a set of streets in a network using a fleet of vehicles with limited capacity located at a central depot. The problem was originally proposed in [3]. The authors suggested a formulation using an undirected connected graph. The use of the terms *arc* and *edge* in the context of Arc Routing Problems is not strict since problem definitions based on directed or undirected graphs can be found in the literature.

The CARP can be stated as follows. Given an undirected connected graph $G(V, E)$, where V denotes the set of nodes and E denotes the set of edges. The node set V contains a special depot node, in which is situated a fleet of identical vehicles of capacity Q. Each edge $e \in E$ has a nonnegative demand d_e and a travel cost c_e. Only edges with $d_e > 0$ are required to be serviced in the solution. Servicing an edge means traversing it in either way. The problem consists in determining a set of routes servicing all required edges, satisfying the total demand while respecting the capacity of vehicles. Even the single vehicle variant called Rural Postman Problem is NP-hard [8].

The Capacitated Arc Routing Problem with Time Windows (CARPTW) extends the CARP. Each required edge is additionally associated with a time window $[f_e, g_e]$ in which the service must begin and a processing time p_e. Arriving earlier than f_e induces a waiting time of the vehicle before the service of e can start. Later arrival than g_e is not allowed. Each edge has a nonnegative traversal time t_e denoting the time needed to traverse the edge without servicing it. The solution of the CARPTW must in addition to the CARP respect the time windows constraints.

The CARPTW practical applications are e.g. the waste collection in urban areas or the winter glittering of road networks. The consideration of time windows can be justified e.g. by specific constraints connected with the traffic rush hour. Peak hours are not suitable for waste collection. Therefore a solution in which the operations are adequately distributed in the day planning horizon is preferable. On the other hand the case of winter glittering suggests that more important roads can have the time window set to

[1]University of Economics, Department of Econometrics, nám. W. Chruchilla 4, 130 67 Prague 3, Czech Republic, jan.melechovsky@vse.cz

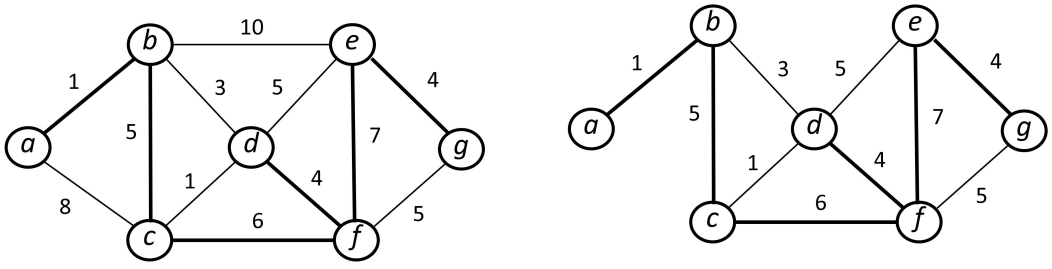
the beginning of the planning horizon while other parts of the road network can wait longer time for the treatment.

The literature dealing with the CARP is larger compared to the CARPTW. Two papers survey the Arc Routing Problems in general [5, 10]. An annotated bibliography of recent solution approaches to different formulations of Arc Routing Problems can be found in [1]. Among heuristic solution approaches developed for the CARP can be mentioned a tabu search algorithm presented in [6]. A heuristic based on partial path extension was proposed in [2].

The CARPTW has received much less attention of the researchers so far. Wøhlk in her dissertation [9] proposed two formulations of the CARPTW (an arc routing formulation and a transformation into a node routing problem) and developed a dynamic programming algorithm combined with a simulated annealing. The author also proposed benchmarks for the CARPTW and designed a column generation to get tight lower bounds. The only published paper dealing with the CARPTW is a book chapter [7]. The authors present a metaheuristic based on the Greedy Randomized Adaptive Search Procedure and a giant tour splitting algorithm. In this approach a solution is encoded as a giant tour visiting all required edges. A splitting algorithm is then used to obtain a set of feasible vehicle routes.

2 The Solution Approach

Before the presentation of the implemented algorithm some additional notation shall be introduced and the problem pre-processing issues are discussed. The set of required edges is denoted R. Some edges $e \notin R$ can be eliminated by a simple preprocessing. An edge $e(i,j)$ can be removed from the graph if there are two edges $u(i,k)$ and $v(k,j)$ such that $c_{ij} \geq c_{ik} + c_{kl}$ and $t_{ij} \geq t_{ik} + t_{kl}$. The elimination is illustrated in Figure 1. The required edges are depicted with bold lines. In this example $c_{ij} = t_{ij}$ for all edges.



(a) An instance of an Arc Routing Problem.

(b) The same instance after the edge elimination procedure.

Figure 1 Illustration of the edge elimination procedure.

Each required edge is then replaced with two opposite arcs a and a' indexed from 0 to $2r-1$, where $r = |R|$. A travel cost matrix C and a travel time matrix T of size $2r \times 2r$ are precomputed. The minimum travel cost distance c_{ab} between arcs $a = (i,j)$ and $b = (k,l)$ is defined as the minimum cost path from j to k. We assume that the travel time is proportional to the travel cost: $t_{ij} = \alpha c_{ij}$, where $\alpha > 0$ is a constant.

2.1 Initial Solution Heuristic

A Path Scanning heuristic (PS) similar to that one proposed in [7] is used to build an initial feasible solution. The method builds successively routes until all required edges are serviced. A step of the algorithm performs an extension of a current route with an unserviced edge. Let w denote the tail node of the last added arc. Let Z denote the set of edges which can be inserted after w, i.e. edges are not serviced yet and the time window constraints and vehicle capacity constraints are respected. Only edges $z \in Z$ with minimal cost between w and z are considered. However, there might be multiple edges with the same cost hence some additional rules are usually applied to break ties. A number of rules has been proposed for the CARP and the CARPTW in the literature. The following rules have been adopted in our implementation:

1. Minimize the cost c_{e0} of the return to the depot,
2. Maximize c_{e0},
3. Use rule 1 if the vehicle is less than half-full, otherwise use rule 2,
4. Minimize the waiting time,
5. Minimize the gap $g_e - a_e$, where a_e is the arrival time to the start of e.

Each rule is used to construct one solution and the best solution is kept.

2.2 The Variable Neighborhood Search

Variable neighborhood search (VNS) is a metaheuristic framework proposed by Mladenovic and Hansen [4]. Its principle idea is to dynamically change neighborhood when a local optimum is reached. The VNS consists of a local search mechanism, a set of different neighborhoods and a set of rules defining the use and the selection of neighborhoods according to search results. In the present implementation of the VNS, the set of neighborhoods $N_k(x)$, $k = 1, \ldots, k_{max}$ of a solution x is defined as the set of all solutions that can be obtained by removing a sequence of k consecutive required arcs from x. The search starts with $k = 1$. Given an incumbent solution x^*, k consecutive required arcs are removed at random from each route of x^*. Let x' be the solution obtained from x^*. The local search, applied to x', tries to refill the routes in x' with yet unsatisfied requests at minimum cost. If the resulting solution x'' dominates x^*, the search continues with x'' as the incumbent solution ($x^* = x''$) and $k = 1$. If not, the neighborhood is enlarged by setting $k = \min(k + 1, k_{max})$ and the search restarts with x^* unchanged. The steps of the VNS are shown in Algorithm 1. The algorithm stops when either the maximum number i_{max} of iterations is encountered or ni_{max} iterations without improvement of the incumbent solution are performed.

The neighborhood definition is depicted in Figure 2. The solution consists of two routes starting and ending in the depot node. Required arcs are in bold and the selected consecutive arcs are enclosed in an ellipsis. The neighborhood size k is dynamically updated during the search and the neighborhood change is assured by the procedure *Perturb* which takes k as a parameter. *Perturb* scans the solution and removes randomly k consecutive required edges from each route. The edges are then reinserted to other position. The feasibility of the solution is preserved.

Algorithm 1: VNS Algorithm

```
1  begin
2      it_count ← 0                                  // Initialize iterations counter
3      ni_count ← 0                                  // Initialize non-improving iterations counter
4      k = k_min                                     // Initialize neighborhood parameter
5      x* ← Path_Scanning_Heuristic                  // Initialize the best solution
6      repeat
7          x' ← Perturb(x*, k)
8          x'' ← Local_Search(x')
9          if Cost(x'') < Cost(x*) then
10             x* ← x''                              // Update the best solution
11             k ← max(k_min, k − 1)                 // Reduce the size of the neighborhood
12             ni_count ← 0                          // Update non-improving iterations counter
13         else
14             k ← min(k_max, k + 1)                 // Enlarge the size of the neighborhood
15             ni_count ← ni_count + 1               // Update non-improving iterations counter
16         end if
17     until it_count = it_max or ni_count = ni_max
18 end
```

The last component of the proposed VNS is the *LocalSearch* procedure. It operates on the solution x' modified by *Perturb*. Its purpose is to find a local optimum in the neighborhood of x'. *LocalSearch* uses two operators: 1) relocate and 2) exchange. Both operators can be applied to a single route or to a pair of routes. Relocate removes and reinserts one or two consecutive edges. Exchange swaps one or

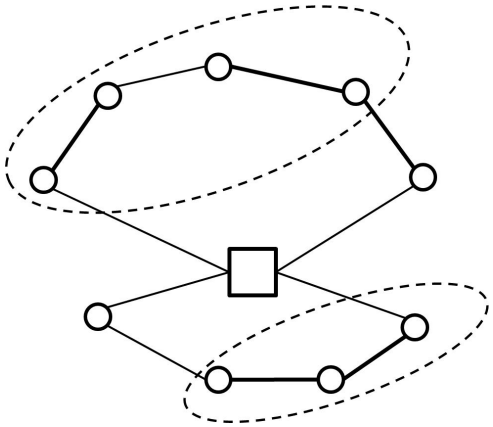


Figure 2 A solution neighborhood example with $k = 2$.

two consecutive edges. Each possible reinsertion is evaluated also with the opposite direction. The two operators are called in a sequence and the first improving move is performed. *LocalSearch* stops when no further improvements are detected.

3 Preliminary Results

The algorithm was implemented in C++ and the experiments were carried out on a 2.7 GHz PC. Preliminary experiments were were executed with the following parameter setting: $it_{max} = 1000$, $ni_{max} = 250$, $k_{min} = 1$, $k_{max} = 5$. As the testing environment served the CARPTW benchmarks proposed in [9]. The benchmarks consist of three sets A, B and C. The total number of edges $|E|$ ranges from 15 to 90 and the number of required edges $|R|$ from 11 to 81. The sets differ mainly in the width of the time windows (tight for A, medium for B and wider for C).

The results of our VNS on the CARPTW instances are presented in Table 1. The first four columns indicate respectively the instance name, the number of nodes, the total number of edges and the number of required edges. BKS is the best known solution value. The entries marked with * present proved optimal values. The results of the VNS are reported in the last two columns. Gap is the percentage difference between the solution value found by the VNS. CPU denotes the computational time in seconds. The presented results are average values obtained over 5 runs of the VNS with a different rand seed setting per run. The results prove a good average performance of the VNS. The algorithm was able to find optimal solutions for most of the small and medium size instances. The average gap over all instances is only 0.5 %. The average computational time 101.3 s is acceptable taking into account the complexity of the CARPTW. Larger computational times are reported for the instances with 40 and 60 nodes. These problems seem to be quite difficult to solve.

4 Conclusions

This paper presented a metaheuristic solution approach based on the VNS for solving the CARPTW. The tackled problem is NP-hard and therefore difficult to solve for real-sized instances. The proposed algorithm relies mainly on a dynamic change of the explored solution neighborhood which enables the algorithm to escape from local optima. The algorithm was tested on a set of CARPTW proposed in the literature. The obtained results indicate a good performance of the VNS.

Further research can be oriented on more complex local search method. An enhancement of the implemented local search with other operators might provide better results. More detailed computational experiments could help to precise the standard parameter setting and optimize the performance with respect to computational time. An evaluation on classical CARP instances might light up the universality of the proposed method.

Instance	$\|V\|$	$\|E\|$	$\|R\|$	BKS	VNS	
					Gap(%)	CPU(s)
A10A	10	15	11	*107	0.0	0.9
A13A	13	23	22	*202	0.0	1.1
A13B	13	23	22	*171	0.2	1.6
A13C	13	23	22	*163	0.1	3.9
A20B	20	31	29	*260	0.0	14.3
A40C	40	69	63	*660	0.8	189.6
A40D	40	69	63	*807	0.9	365.9
A60A	60	90	81	*1822	1.1	412.4
B10A	10	15	11	*87	0.0	1.3
B13A	13	23	22	*167	0.0	2.9
B13B	13	23	22	*152	0.0	3.8
B13C	13	23	22	*141	0.0	5.6
B20B	20	31	29	*214	0.1	4.7
B40C	40	69	63	602	0.6	178.3
B40D	40	69	63	*730	0.7	258.9
B60A	60	90	81	1554	1.5	298.4
C10A	10	15	11	*73	0.0	1.5
C13A	13	23	22	*142	0.0	0.9
C13B	13	23	22	*132	0.0	2.6
C13C	13	23	22	*121	0.0	4.3
C20B	20	31	29	*186	0.1	6.2
C40C	40	69	63	545	0.8	145.3
C40D	40	69	63	626	2.9	201.3
C60A	60	90	81	*1283	1.8	325.9
Average					0.5	101.3

Table 1 The performance of the VNS on CARPTW instances.

Acknowledgements

The research was fully supported by the Czech Science Foundation, project reg. number P402/12/P635. This financial support is kindly acknowledged.

References

[1] Corberán, A., & Prins, C. Recent results on arc routing problems: An annotated bibliography. *Networks* **56**, (2010), 50–69.

[2] Fábry, J., & Pelikán, J. Capacitated postman problem. In *Mathematical Methods in Economics 2013*, College of Polytechnics Jihlava, 2013, 153–158.

[3] Golden, B. L., & Wong, R. T. Capacitated arc routing problems. *Networks* **11**, (1981), 305–315.

[4] Hansen, P., & Mladenović, N. Variable neighborhood search. In E. Burke, & G. Kendall (Eds.), *Search Methodologies*, Springer US, 2005, 211–238.

[5] Hertz, A. Recent trends in arc routing. In M. C. Golumbic, & I. B.-A. Hartman (Eds.), *Graph Theory, Combinatorics and Algorithms*, Springer US volume 34 of *Operations Research/Computer Science Interfaces Series*, 2005, 215–236.

[6] Hertz, A., Laporte, G., & Mittaz, M. A tabu search heuristic for the capacitated arc routing problem. *Journal of Parallel and Distributed Computing* **48**, (2000), 129 – 135.

[7] Labadi, N., Prins, C., & Reghioui, M. Grasp with path relinking for the capacitated arc routing problem with time windows. In A. Fink, & F. Rothlauf (Eds.), *Advances in Computational Intelligence in Transport, Logistics, and Supply Chain Management*, Springer Berlin Heidelberg volume 144 of *Studies in Computational Intelligence*, 2008, 111–135.

[8] Lenstra, J. K., & Kan, A. H. G. R. On general routing problems. *Networks* **6** (1976), 273–280.

[9] Wøhlk, S. *Contributions to Arc Routing*. Ph.D. thesis University of Southern Denmark, 2006.

[10] Wøhlk, S. A decade of capacitated arc routing. In B. Golden, S. Raghavan, & E. Wasil (Eds.), *The Vehicle Routing Problem: Latest Advances and New Challenges*, Springer US volume 43 of *Operations Research/Computer Science Interfaces Series*, 2008, 29–48.

Credit risk in the Czech construction Sector

Aleš Melecký[1]

Abstract. The global recession has increased credit risk in the Czech banking system as the macroeconomic fundamentals weakened. This led to the reduction of the public spending on construction works. The paper analyzes the sensitiveness of credit risk in the Czech construction sector, measured by the ratio of non-performing loans to total loans (NPLR) granted in this sector by the banking institutions, to changes in macroeconomic fundamentals. The author uses quarterly data from 2002 to 2013 and estimates the sensitivity of the NPLR to unemployment rate, PPI inflation, lending interest rates, and the change of real effective exchange rate of the Czech Crown (CZK). The author employs a vector error correction model (VECM) which is suitable for cointegrated series and can distinguish long run relations and short run dynamics. The author finds that, macroeconomic variables are significant determinants of NPLR in Construction sector. VECM reveals high and statistically significant impact of the change in the companies´ lending rate and unemployment rate to the change of NPLR in the Czech Republic. These findings may help policy-makers to better understand behavior of NPLR in this endangered sector and improve their reaction in order to prevent possible spillovers to the entire economy.

Keywords: Credit risk, Non-performing loans ratio, Czech Republic, Construction sector, Vector error correction model.

JEL Classification: C32, E32, G28
AMS Classification: 62M10

1 Introduction

Non-performing loans[2] (NPLs) are broadly used in the empirical literature as an indicator of credit risk materialization, see e.g. Buncic and Melecky [3], Louzis et al. [7] or Festic et al. [4]. Increase in credit risk can limit credit supply and reduce investment activity and as a result constrain economic activity. The aim of this paper is to analyze the sensitivity of the credit risk in the Czech construction sector, measured by the non-performing loans ratio (NPLR) to changes in macroeconomic fundamentals.[3]

This paper focuses on the Czech construction sector which undergone dynamic development during past 12 years. In 2002 most of the construction works were financed by the public sector (55.1%), but the share of public spending decreased over time (with the exception of 2009). In 2012, majority of the projects were financed by the private contractors and public share on the construction works was reduced to only 38.3%, see figure 1.

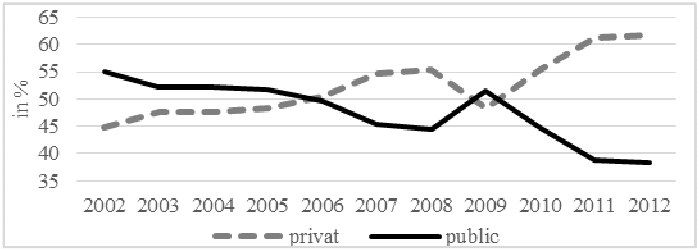


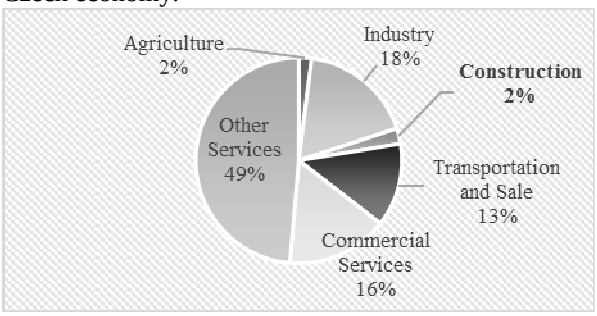
Figure 1 Private and public shares on construction works
Source: Self-elaboration based on the Czech National Banks' data.

[1] VSB - Technical University of Ostrava, Faculty of Economics, Department of Economics, Sokolska tr. 33, 701 21 Ostrava 1, Czech Republic, e-mail: ales.melecky@vsb.cz

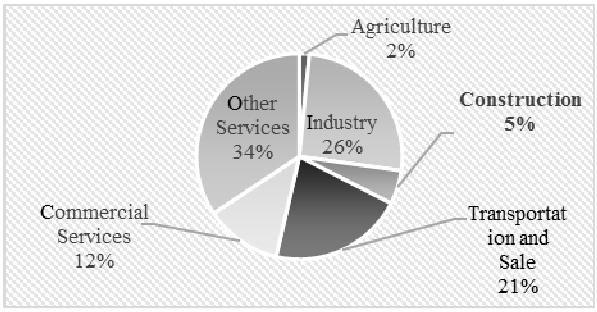
[2] The Czech National Bank follows the broadly accepted definition of NPLs published by the IMF in the Financial Soundness Indicators Compilation Guide of March 2006, see IMF [6]. Loans should be classified as NPL when (1) payments of principal and interest are past due by three months (90 days) or more, or (2) interest payments equal to three months (90 days) interest or more have been capitalized (re-invested into the principal amount), refinanced, or rolled over (i.e. payment has been delayed by arrangement).

[3] Non-performing loans ratio is calculated as non-performing loans in the Construction sector to total loans granted in this sector by the banking institutions. The data includes information about non-financial companies only.

The situation of Construction sector is specific within the Czech economy. Figure 2 shows average shares[4] of loans in respective sector as part of total loans granted in the 6 sectors of economy (Panel A) and similarly average share of NPLs in each sector (Panel B) in the Czech Republic. Based on the graph in the Panel A we can conclude that Construction sector is relatively small (according to the total granted loans) compared to the other sectors (with only 2% share) and it is approximately the same size as Agriculture sector. In the panel B we can see that share of NPLs in Construction sector on total NPLs is on average 2.5 times higher than in Agriculture sector. According to calculated averages, the NPLR in Construction sector is the highest of all sectors in the Czech economy.



Panel A Average share of loans in sectors as part of total loans in 6 sectors (2002-2013)



Panel B Average share of NPLs in sectors as part of total NPLs in 6 sectors (2002-2013)

Figure 2 Granted loans and non-performing loans in sectors (2002-2013)
Source: Self-elaboration based on Czech National Banks' data.

Dynamic development of NPLRs at the aggregate level, in Construction sector and in Agriculture sector is captured in figure 3. As can be seen in the figure, the dynamics of NPLR in Construction sector varies significantly from aggregate level as well as from Agriculture sector. The big differences among time series of NPLRs in the beginning of 2002 were quickly reduced and the development of the NPLRs in Construction sector was quite stable up to 2008 when the Czech Republic was fully hit by the impacts of the global financial crisis. Between the years 2008 and 2013 reduction of the public spending, decrease of property prices and overall uncertainty in the economy caused significant increase in the NPLR in the Czech construction sector.

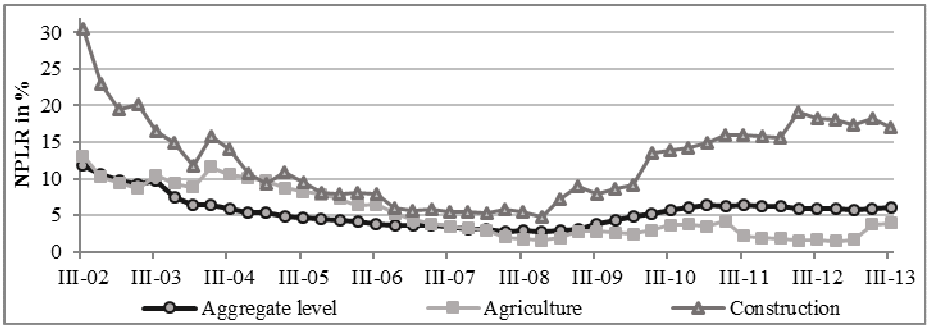


Figure 3 NPLR at aggregate level, in Agriculture sector and in Construction sector (2002-2013)
Source: Self-elaboration based on CNB's data.

Boss et al. [2] showed that the impact and sometimes even signs of effect of the macroeconomic factors on the credit risk could differ significantly across sectors of modeled economy. Also Melecky, Sulganova, Lelkova [8] confirmed different effects of macroeconomic variables on the NPLRs across sectors (and in comparison with aggregate level) with the use of ARDL approach. It is therefore appropriate to employ more sophisticated method of estimation (such as vector error correction model) to perform detailed analysis of the Czech construction sector. This sector could be potential source of future problems for entire economy if the unpleasant development will continue.

2 Data and Methodology

2.1 Data

The empirical analysis is based on the quarterly data for the Czech Republic from the first quarter of 2002 to the end of 2013 which represent a maximum data span available to public. Primary characteristics of the data and

[4] Presented averages are calculated as 12-year arithmetic averages in 2002 to 2013 period.

their sources are presented in table 1. Potential credit risk is measured by non-performing loans ratio in the Czech construction sector. For the pricing of lending in Construction sector author uses companies´ lending rate. Exchange rate dynamics is captured by annual growth rate of real effective exchange rate of CZK as suggested by Mojon and Peersman [9] or Babouček and Jančar [1]. Unemployment rate was found to be significant determinant of non-performing loans in a lot of papers, see Babouček and Jančar [1] or Festić and Bekő [5] among others. Increase in unemployment rate leads to decrease of income, reduction of investment and rise of NPLR. The author tested also real GDP growth in Construction sector and at the aggregate level but they were not significant and caused instability of the model. Table 2 provides basic descriptive statistics of the data used in the empirical analysis. Time series were obtained from the Czech National Banks' database ARAD and the Czech Statistical Office.

Variable	Characteristic	Source
NPLs ratio - Construction	Calculated as a share of non-performing loans in the Construction sector to total volume of nominal loans in this sector	ARAD, CNB
Lending rate - Companies	Nominal lending rate for companies	ARAD, CNB
Real effective exchange rate growth	Annual growth of the real effective exchange rate of the Czech Crown	ARAD, CNB
Inflation	Calculated as annual percentage change of the PPI index (2005=100)	ARAD, CNB
Unemployment rate	General unemployment rate of the aged 15 to 64 years	CSO

Table 1 List of variables and their sources

Abbreviation	Variable	Mean	Median	Max.	Min.	Std. Dev.
NPLR_C	Non-performing loans ratio - Construction	12.74	13.64	30.60	4.79	5.77
INFL	Inflation	2.05	1.73	7.85	-5.43	2.97
LRC	Lending rates - Companies	4.43	4.31	6.32	3.07	0.74
REER	Real effective exchange rate growth	1.81	0.46	14.59	-7.82	5.33
U	Unemployment rate	6.95	7.15	8.40	4.30	1.10

Table 2 Descriptive statistics

2.2 Methodology

In this paper I use vector error correction model (VECM) to estimate effects of the macroeconomic determinants on NPLR. Especially I focus on short run dynamics described by the error correction model. VECM can be generally characterized as in equation 1:

$$\Delta Y_t = \Pi_0 Y_{t-1} + \sum_{j=1}^{p} B_{0,j} \Delta Y_{t-j} + u_t \,, \tag{1}$$

where $\Delta Y_t = Y_t - Y_{t-1}$, Y_t is an m-dimensional vector, $\Pi_0 = \alpha_0 \beta_0'$ has rank $0 \le r_0 \le m$, $B_{0,j}$ $(j = 1,..., p)$ are $m \times m$ (transient) coefficient matrices and u_t is an m-vector error term with mean zero and nonsingular covariance matrix $\sum_u$. The rank r_0 of Π_0 is an order parameter and measure the cointegrating rank or the number of cointegrating relations in the system, for more details see e.g. Zhipeng and Phillips [10]. The vector of variables includes non-performing loans ratio in Construction sector, companies' lending rate, inflation, real effective exchange rate growth, and unemployment rate. The lag order p characterize the transient dynamics in the system. However Johansen cointegrating test indicated up to 3 cointegrating vectors, the loss of explanatory power of the model in the case of using 1 cointegrating vector is very small and the economic interpretations of the long run relations are then relatively straightforward. Therefore, I prefer and present results of this specification in the following section of the paper.

3 Estimation results

With the use of the data described above I estimated following representation of the VECM (see table 3). In the first part of table 3 one can find cointegrating vector and the second part provides estimation of the error correction part including most important information about the explanatory power of the model[5].

[5] Other results are available upon request. The model was tested for autocorrelation and normality of residuals as well as the stability based on the inverse root test.

Cointegrating equation					
C	**NPLR_C(-1)**	**INFL(-1)**	**LRC(-1)**	**REER(-1)**	**U(-1)**
-165.6801	1.000000	0.953012	20.87840	-3.324161	9.536118
		(0.98048)	(5.11542)	(1.02152)	(1.10667)
		[0.97199]	[4.08146]	[-3.25413]	[8.61694]
Short-term dynamics of the vector error correction model					
Error Correction:	**D(NPLR_C)**	**D(INFL)**	**D(LRC)**	**D(REER)**	**D(U)**
CointEq1	-0.265165	-0.058466	-0.010981	0.072836	-0.006233
	(0.03729)	(0.05706)	(0.00813)	(0.12792)	(0.01153)
	[-7.11173]	[-1.02469]	[-1.35045]	[0.56940]	[-0.54068]
D(NPLR_C(-1))	-0.427629	-0.174551	-0.076241	-0.660578	0.030912
	(0.14012)	(0.21443)	(0.03056)	(0.48072)	(0.04333)
	[-3.05181]	[-0.81403]	[-2.49489]	[-1.37413]	[0.71346]
D(NPLR_C(-2))	-0.543729	-0.026294	0.012553	-0.326070	-0.028658
	(0.13873)	(0.21229)	(0.03025)	(0.47593)	(0.04289)
	[-3.91942]	[-0.12386]	[0.41491]	[-0.68512]	[-0.66810]
D(NPLR_C(-3))	-0.164281	0.320867	-0.017506	0.454009	-0.003981
	(0.13031)	(0.19941)	(0.02842)	(0.44705)	(0.04029)
	[-1.26073]	[1.60912]	[-0.61603]	[1.01558]	[-0.09881]
D(NPLR_C(-4))	0.488926	-0.045121	-0.001040	0.583238	-0.028079
	(0.12396)	(0.18969)	(0.02703)	(0.42526)	(0.03833)
	[3.94437]	[-0.23787]	[-0.03848]	[1.37149]	[-0.73263]
D(NPLR_C(-5))	0.314696	-0.008559	0.017325	0.079704	0.002933
	(0.12623)	(0.19317)	(0.02753)	(0.43306)	(0.03903)
	[2.49302]	[-0.04431]	[0.62934]	[0.18405]	[0.07514]
D(INFL(-1))	0.387709	0.075695	0.084258	0.508893	0.087514
	(0.16796)	(0.25703)	(0.03663)	(0.57624)	(0.05193)
	[2.30829]	[0.29449]	[2.30022]	[0.88313]	[1.68508]
D(INFL(-2))	-0.193486	0.304635	0.006127	-0.005752	0.141915
	(0.15265)	(0.23360)	(0.03329)	(0.52370)	(0.04720)
	[-1.26751]	[1.30410]	[0.18405]	[-0.01098]	[3.00670]
D(INFL(-3))	0.393140	0.269434	-0.014986	-0.104871	-0.004674
	(0.13727)	(0.21007)	(0.02994)	(0.47095)	(0.04244)
	[2.86393]	[1.28262]	[-0.50058]	[-0.22268]	[-0.11011]
D(INFL(-4))	0.501335	-0.079567	0.036599	0.412698	0.014896
	(0.15072)	(0.23064)	(0.03287)	(0.51708)	(0.04660)
	[3.32626]	[-0.34498]	[1.11344]	[0.79813]	[0.31964]
D(INFL(-5))	0.563688	0.595199	0.042104	0.779411	0.019524
	(0.14815)	(0.22671)	(0.03231)	(0.50827)	(0.04581)
	[3.80479]	[2.62532]	[1.30315]	[1.53346]	[0.42622]
D(LRC(-1))	5.265583	1.341862	0.535600	2.998868	-0.739672
	(1.49050)	(2.28089)	(0.32506)	(5.11351)	(0.46086)
	[3.53276]	[0.58831]	[1.64772]	[0.58646]	[-1.60497]
D(LRC(-2))	4.153809	3.813642	0.115653	9.384200	0.320124
	(1.52430)	(2.33261)	(0.33243)	(5.22945)	(0.47131)
	[2.72507]	[1.63493]	[0.34791]	[1.79449]	[0.67922]
D(LRC(-3))	1.248883	-1.801444	-0.161619	1.116207	0.205377
	(1.15908)	(1.77372)	(0.25278)	(3.97649)	(0.35839)
	[1.07748]	[-1.01563]	[-0.63937]	[0.28070]	[0.57306]
D(LRC(-4))	-2.025506	-1.970304	0.263096	-0.921729	-0.272320
	(1.21090)	(1.85303)	(0.26408)	(4.15428)	(0.37441)
	[-1.67272]	[-1.06329]	[0.99628]	[-0.22187]	[-0.72733]
D(LRC(-5))	-2.674173	2.332007	0.066312	8.309636	0.865580
	(1.51304)	(2.31538)	(0.32997)	(5.19082)	(0.46783)
	[-1.76742]	[1.00718]	[0.20096]	[1.60083]	[1.85019]
D(REER(-1))	-0.847136	-0.241835	-0.047422	-0.330691	-0.054992
	(0.12678)	(0.19401)	(0.02765)	(0.43494)	(0.03920)
	[-6.68205]	[-1.24653]	[-1.71519]	[-0.76031]	[-1.40287]
D(REER(-2))	-0.438461	-0.245559	-0.039061	-0.258990	-0.065329
	(0.11197)	(0.17134)	(0.02442)	(0.38412)	(0.03462)

	[-3.91603]	[-1.43317]	[-1.59970]	[-0.67424]	[-1.88703]
D(REER(-3))	-0.424885	-0.411530	-0.028383	-0.427621	-0.024565
	(0.10453)	(0.15996)	(0.02280)	(0.35861)	(0.03232)
	[-4.06481]	[-2.57275]	[-1.24509]	[-1.19245]	[-0.76005]
D(REER(-4))	-0.315657	-0.300454	-0.033265	-0.755232	0.029701
	(0.08741)	(0.13377)	(0.01906)	(0.29990)	(0.02703)
	[-3.61102]	[-2.24606]	[-1.74495]	[-2.51831]	[1.09886]
D(REER(-5))	-0.263807	-0.289596	-0.014497	-0.370876	-0.011894
	(0.08074)	(0.12355)	(0.01761)	(0.27699)	(0.02496)
	[-3.26749]	[-2.34394]	[-0.82333]	[-1.33897]	[-0.47647]
D(U(-1))	2.831322	-0.766526	-0.109735	-1.590425	0.243844
	(0.64032)	(0.97988)	(0.13964)	(2.19677)	(0.19799)
	[4.42171]	[-0.78227]	[-0.78582]	[-0.72398]	[1.23161]
D(U(-2))	0.918339	-1.846820	0.005126	-1.292019	0.374870
	(0.80963)	(1.23896)	(0.17657)	(2.77761)	(0.25034)
	[1.13427]	[-1.49062]	[0.02903]	[-0.46516]	[1.49746]
D(U(-3))	2.755872	2.178413	-0.126523	-0.334947	0.542110
	(0.77832)	(1.19105)	(0.16974)	(2.67019)	(0.24066)
	[3.54081]	[1.82899]	[-0.74540]	[-0.12544]	[2.25264]
D(U(-4))	2.579860	-0.049447	0.321680	1.179548	-0.153993
	(0.96154)	(1.47142)	(0.20970)	(3.29877)	(0.29731)
	[2.68306]	[-0.03361]	[1.53403]	[0.35757]	[-0.51796]
D(U(-5))	3.515119	3.083695	0.187530	4.357153	-0.226570
	(1.04199)	(1.59454)	(0.22724)	(3.57479)	(0.32218)
	[3.37347]	[1.93391]	[0.82524]	[1.21886]	[-0.70323]
C	0.004581	-0.209222	-0.030944	0.449737	-0.007132
	(0.16951)	(0.25940)	(0.03697)	(0.58154)	(0.05241)
	[0.02703]	[-0.80657]	[-0.83707]	[0.77336]	[-0.13608]
Adj. R-squared	0.789542	0.529043	0.278750	0.473833	0.564887
Akaike AIC	2.709986	3.560894	-0.335755	5.175531	0.362454
Schwarz SC	3.827060	4.677967	0.781318	6.292605	1.479528
Log likelihood		-81.92867			
Akaike information criterion		10.56803			
Schwarz criterion		16.36026			

Table 3 Estimation results - VECM

Because the estimation results include a lot of information and the paper size is limited, the author focuses only on the most relevant ones with respect to the goal of the paper id est. how are differences in the NPLR influenced by the differences in the main macroeconomic determinants. The obtained t-statistics enable to calculate the p-values to judge the statistical significance of macroeconomic determinants. Estimation results of error correction model suggest that all macroeconomic drivers of NPLR are statistically significant at 5%. According to the lag exclusion test none of the lags can be excluded from the model.

The estimation results suggest that companies´ lending rate and unemployment rate are the most important macroeconomic determinants of the short-run dynamics of NPLR (according to the absolute values of estimated coefficients) in the Construction sector. In the case of the effect of companies´ lending rate on difference in NPLR, the estimated coefficients are statistically significant on first two lags, and their estimated coefficients are the highest of all. These results suggest that an increase in interest rate is associated with higher cost of lending for the debtors and lead to an increase in NPLR. Statistically significant positive values of estimated coefficients suggest that higher inflation leads to an increase in NPLR. This might be a result of worsening expectation about the future, see Festić and Bekő [5] or negative impact of unanticipated inflation on banking sector and worsening of debt recoveries. Estimated negative values of coefficients of the change in the real exchange rate imply that faster depreciation of the Czech Crown could be a positive impulse for the Czech construction sector and can decrease the rise of NPLR. In the case of differences in unemployment rate, I found its positive relation to the difference in NPLR. This result is in line with economic theory when the lower unemployment leads to a faster decline of NPLR due to better financial position of economic subjects, improved ability to repay their debts, and supports long run investment activity.

4 Conclusion

The Czech construction sector undergone dynamic development during the past 12 years and dynamics of non-performing loans in this sector is significantly different from the dynamics at the aggregate level as well as in other sectors of the Czech economy. Prior to the global financial crisis the situation in this sector was quite stable but between the years 2008 and 2013 the reduction of the public spending, decrease of property prices and overall uncertainty in the Czech economy caused significant increase of the NPLR in the Construction sector.

The aim of this paper was to analyze the sensitivity of the credit risk, measured by the non-performing loans ratio (NPLR), to macroeconomic environment in the Czech Construction sector. With the use of vector error correction model I identified companies´ lending rates, inflation calculated from PPI index, real effective exchange rate of the Czech Crown and unemployment rate as significant macroeconomic determinants for the dynamics of non-performing loans ratio in the Czech construction sector. Based on the error correction model, which examined relations between variables at differences, increase in real effective exchange rate lead to the reduction of credit risk measured by NPLR. On the other hand, rise in companies´ lending rate, inflation and unemployment rate increases credit risk in Construction sector.

Construction sector had on the average the highest level of NPLR across examined sectors in the Czech Republic during 2002-2013 period. NPLR in the Czech construction sector were increasing sharply from the beginning of 2008 due to the effects of the global financial crisis. In the light of this undesirable development, discussion about the public spending plan in this sector should be reopened. Construction sector should be closely watched by the authorities in order to prevent potential spillovers to the entire economy.

Acknowledgements

This paper has been elaborated in the framework of the project No. CZ.1.07/2.3.00/30.0016 supported by Operational Programme Education for Competitiveness and co-financed by the European Social Fund, and the Czech Science Foundation grant No.13-20613S.

References

[1] Babouček, I., and Jančar, M.: A VAR analysis of the effects of macroeconomic shocks to the quality of the aggregate loan portfolio of the Czech banking sector. *Czech National Bank Working Paper Series No. 1/2005.* [online]. 2005. [cit. 21. 4. 2014]. Available at: http://www.cnb.cz/en/research/research_publications/cnb_wp/download/cnbwp_2005_01.pdf

[2] Boss, M., Fenz, G., Pann, J., Puhr, C., Schneider, M., and Ubl, E.: Modeling Credit Risk through the Austrian Business Cycle: An Update of the OeNB Model. *Financial Stability Report 17*. [online]. 2009. [cit. 21. 4. 2014]. Available at: http://www.oenb.at/en/Publications/Financial-Market/Financial-Stability-Report/2009/Financial-Stability-Report-17.html

[3] Buncic, D., and Melecky, M.: Macroprudential stress testing of credit risk: a practical approach for policy makers. *Journal of Financial Stability* **9** (2013), 347-370.

[4] Festić, M., Kavkler, A., and Repina, S.: The macroeconomic sources of systemic risk in the banking sectors of five new EU member states. *Journal of Finance and Banking* **35** (2011), 310 – 322.

[5] Festić, M., and Bekő, J.: The banking sector and macroeconomic performance in central european economies. *Czech Journal of Economics and Finance* **58** (2008), 131 – 151.

[6] IMF: *Financial Soundness Indicators Compilation Guide*. IMF. [online]. 2006. [cit. 21. 4. 2014]. Available at: http://www.imf.org/external/pubs/ft/fsi/guide/2006/index.htm

[7] Louzis, D. P., Vouldis, A. T., and Metaxas, V. L.: Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. *Journal of Finance and Banking* **36** (2012), 1012 – 1027.

[8] Melecky, A., Sulganova, M., and Lelkova, P.: Credit Risk and the Czech Macroeconomy: The Differences between Aggregate and Sectoral Levels. *Proceedings of the 14th International Conference on Finance and Banking* : 16-17 October 2013, Ostrava, Czech Republic, Silesian University, School of Business Administration, 2013, s. 267-276. 978-80-7248-939-8.

[9] Mojon, B., and Peersman, G.: A VAR description of the effects of monetary policy in the individual countries of the Euroarea. *ECB Working Paper Series No. 92*. [online]. 2001. [cit. 21. 4. 2014]. Available at: https://biblio.ugent.be/input/download?func=downloadFile&recordOId=145330&fileOId=732902

[10] Zhipeng, L., and Phillips, P. C. B.: Automated Estimation of Vector Error Correction Models. *Cowles Foundation Discussion Paper No. 1873.* [online]. 2010. [cit. 21. 4. 2014]. Available at: http://www.econ.ucla.edu/people/papers/Liao/Liao547.pdf

A Cost-based Model for Support of Decision-making about a Spare-parts Storage

Pavla Melicheríková[1]

Abstract. Results of a management decision, whether to establish or not to establish a spare-parts stock, reflect in company costs. This decision problem is often omitted by some companies which operate on a market of building constructions. Therefore, I came out with the mathematical model for the evaluation of spare-parts storage costs.

The calculation of the spare-parts storage costs can be done in two ways. The first one is the analytical calculation based on costs of shortage and superfluous inventory units. The second possibility is to use Monte Carlo simulation. The second way of the calculation will be described in detail because the first one is based on the well-known inventory model. Monte Carlo calculation can be done in four variations which will be explained. After that, Monte Carlo calculation and the interpretation of the results will be presented in an example.

Moreover, some possible decision-making rules will be suggested. Some interesting features of the model will be pointed out. At the end, some possibilities of further research in this field will be stated.

Keywords: cost-based calculation, decision-making support, Monte Carlo simulation, spare-parts storage.

JEL classification: G31 Fixed Investment and Inventory Studies
AMS classification: 90B05 Inventory, storage, reservoirs

1 Fundamentals of the cost-based model

The model was originally proposed for a company which built an equipment complex. The model is purposed to allow company management to decide if establishing of the equipment complex spare-parts storage is a good idea. Some problems of spare-parts are described in [5] and in related references. Some inventory problems are solved in [3]. However, I have not found any suitable model for my problem. Therefore, I have devised the model which will be introduced.

1.1 Assumptions

Every mathematical model is a simplified and abstract picture of a reality. This reality picture should keep all important elements and links between them. For the simplification, I have done three assumptions: choice of an appropriate type of the model, choice of a suitable probability distribution of the spare-parts demand and determination of a modelled time period.

The model is based on the idea of *stochastic inventory model*. Some of these models can be found in [4] and in [3]. Other interesting ideas and recommendations for the model creation are stated in [2].

For Monte Carlo simulation, it is necessary to determine the suitable probability distribution. I have decided to use a *triangular probability distribution* which is good for modelling of random quantities based on a little information about its distribution. Triangular distribution holds three parameters: a, b, and c. In the model, a denotes a minimum spare-part demand, b denotes a maximum spare-part demand and c denotes an expected spare-part demand. These parameters have to be estimated by experts. It is necessary to estimate all parameters of each spare-parts type by each expert.

[1]Palacký University Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, pavla.melicherikova01@upol.cz

The time period must be kept in mind when the model is designed. I have decided to consider *only one spare-parts supply during the field test* of the equipment complex because transport costs of all items of majority spare-parts types are usually the same. It means that each item of these spare-parts needs an individual transportation. Other reason to choose this time period were the highest costs calculated based on the estimated demands. In this paper, spare-parts demand and order of spare-parts are considered to be identical because only one supply of spare-parts is taken into account.

Type of model, triangular distribution of the spare-parts demand and selected time period affect the model results. Moreover, there will be discussed another factors which can influenced the model, however, it is necessary to introduce a demand calculation before it.

1.2 Spare-parts demand calculation

At the beginning of the calculation, it is important to know a spare-parts demand. The demand value can be achieved in two ways: via costs of shortage and superfluous inventory units and via Monte Carlo simulation.

The first possibility of the demand calculation was originally published in [4]. It is based on knowledge of a shortage inventory unit cost and on a unit cost of the superfluous inventory. However, the information about each storage type need not be easy to obtain. I also did not get them from the company. Therefore, results gained by this calculation will not be shown in an example. For more information, see [4] or [7].

The second way, how to obtain a demand estimate, is to use Monte Carlo simulation. Monte Carlo simulation is described in [1]. For this calculation, it is necessary to obtain parameters a, b and c of triangular distribution of spare-parts types from a few experts as was written in the previous subsection. A computer generator generates pseudo-random numbers r_i, $i = 1, 2, \ldots, n$, where n denotes a number of calculation repetitions. Based on parameters of triangular distribution, a probability p is calculated. After that, the spare-parts type demand d is calculated using inverse transformation of the distribution function. The calculations is done using following formulas:

$$p = \frac{c-a}{b-a},$$

$$d = a + \sqrt{r\,(b-a)\,(c-a)} \qquad \text{if } r < p, \tag{1}$$

$$d = b - \sqrt{(1-r)\,(b-a)\,(b-c)} \qquad \text{if } r \geq p. \tag{2}$$

A derivation of inverse transformation formulas is well-known, e.g. it is shown in [6] and in [7]. In every calculation repetition, a computer gains a new pseudo-random number. Based on the number, the computer calculates demand of spare-parts type. Next, this demand is used to calculation of costs which is described in the next section. This calculation is done n-times for every spare-parts type. This is fundamental principle of Monte Carlo. For more informations about Monte Carlo, see [1].

1.3 Spare-parts cost calculation

We need to calculate an appropriate costs index to support the decision-making. I have called it *decision costs*. First, we have to determine a *capital tied in inventories* c_I and *storage costs* c_S. We calculate *total costs* c_T and *opportunity costs* c_O based on these costs. *Decision costs* c_D, which should be compared with another company indicator to support the right decision, are gained as last.

An unit price of a spare-parts type is denoted by c_P and an unit storage cost of a spare-parts type is denoted by c_1. The need of the type of spare-parts, i.e. the spare-parts type demand, is denoted by d as in the previous section. The capital tied in inventories c_I and storage costs c_S have to be calculated for each type of spare-parts. Resultant capital tied in inventories c_{II} and storage costs c_{SS} is gained by *summing up* the costs of all spare-parts types. The type costs are calculated by following formulas:

$$c_I = \frac{1}{2} d \cdot c_P,$$

$$c_S = \frac{1}{2} d \cdot c_1.$$

An interest rate of an alternative investment, denoted i, is the interest rate which the company would gain, if the company invested free money, e.g. in bonds, or saved them on a savings account. Total costs c_T, opportunity costs c_O and decision costs c_D are calculated based on following equations:

$$\begin{aligned} c_T &= c_{II} + c_{SS}, \\ c_O &= c_T \cdot i, \\ c_D &= c_{SS} + c_O. \end{aligned}$$

Total costs, opportunity costs and decision costs can be used to gain overview about costs of the contemplated spare-parts stock. Decision costs are costs which the company should take into account, if they decided to make the spare-parts stock instead of ordering each spare-part when it is needed. In the next section, it is stated how to calculate average costs of all experts.

1.4 Spare-parts average cost calculation

For the calculation, the parameters of triangular probability distribution have to be estimated by experts as explained in section 1.1. The calculation of the spare-parts demand and costs, which was described above, can be done for every expert separately. To gain decision costs for all experts together, it is not correct to use arithmetic average because triangular distributions are not invariant under a linear transformation. One of the possible ways to calculate average values is to use a *weighted average over the variances* of triangular distributions.

Let us assume that we gained parameters estimations from k experts for each type of spare-parts. The variance of the spare-parts type demand for each expert $i = 1, \ldots, k$ and the parameters of triangular distribution should be calculated by following:

$$\sigma_i^2 = \frac{a_i^2 + b_i^2 + c_i^2 - a_i b_i - a_i c_i - b_i c_i}{18}, \tag{3}$$

$$a_{avg} = \frac{\sum_{i=1}^{k} \frac{a_i}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}}, \quad b_{avg} = \frac{\sum_{i=1}^{k} \frac{b_i}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}} \text{ and } c_{avg} = \frac{\sum_{i=1}^{k} \frac{c_i}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}}.$$

Resultant parameters a_{avg}, b_{avg} and c_{avg} have to be calculated for each spare-parts type. Then, for each spare-parts type do: $a = a_{avg}$, $b = b_{avg}$, $c = c_{avg}$ and estimate the spare-parts type demand as stated in section 1.2. Based on the demand, it is possible to calculate costs using formulas in section 1.3. However, there can be one problem with calculation of average parameters which is described and solved in the next section.

1.5 Variants of Monte Carlo simulation

Whole calculation of cost variables described in sections 1.2 and 1.3 should be stochastic. However, computers generate pseudo-random numbers. So, pseudo-random numbers generated by some older generators can fail tests of randomness. Another factors, which can effect the results, are rounding of estimated demand and approach to gain average parameters if a special case of expert parameters estimation occurs.

The spare-parts type demand is not an integer, if it is calculated using Monte Carlo simulation, i.e. by formulas (1) or (2). It is impossible to order a part of some items. Therefore, handling with the estimated demand have consequences in cost results. Is it possible to use *two ways of dealing with the demand*. First one is to round it mathematically. It is the first thing which a lot of people think about. However, is it the correct approach? From the another point of view, if the calculated demand is between 1 and 2, it means that we need 2 pieces because 1 is not enough. So, the demand is always rounded up if the second method is used.

Some expert can estimate parameters of spare-parts type triangular distribution as one exact number. It means that *all estimated parameters of one type of spare-parts will be the same*, i.e. $a = b = c$. For the demand calculation based on the estimation of this expert, it is not any problem. However, if we calculated the variance of this triangular distribution by (3), we should gain zero. Then, it is impossible

to calculate average parameters of the distribution because we cannot divide by zero. There are also two possible ways to get solution. Using the first one, the exact parameters estimate is taken as the best because its weight $\frac{1}{0} = \infty$ is infinitely high. The second way is a small modification of the minimal value a and the maximal value b. It means to add and to subtract a very small number to a and from b, respectively. After this modification, the variance is not zero. Therefore, it is possible to use formulas from the previous section to calculate average parameters.

Based on the choice of the rounding way and the approach to the exact parameters estimation, we can distinguish four variants of the calculation. However, dealing with the exact parameters estimations should influence only the cost calculation based on average parameters. In the next section, this is also shown.

2 Example

From a representative of the company, I have received a dataset containing triangular distribution parameters estimations of 47 spare-parts types. The original data are stated in an appendix of [7]. For the calculation of all costs which are described in the previous section, it is necessary to determine the interest rate of the alternative investment. As alternative investment, I have chosen a business savings account. I have looked up the savings account interest rates of some banks. As the interest rate of the alternative investment, I have chosen the one which was the closest to a half of interest rates span. Than, I have calculated cost for each expert and also the average cost for all experts.

The calculation was done for all Monte Carlo simulation variants. In the tables with results, variants are denoted:

- *Variant 1* - the demand is rounded up and the same parameters of the triangular distribution are modified.
- *Variant 2* - the demand is rounded up and the same parameters of the triangular distribution are taken as the best.
- *Variant 3* - the demand is rounded mathematically and the same parameters of the triangular distribution are modified.
- *Variant 4* - the demand is rounded mathematically and the same parameters of the triangular distribution are taken as the best.

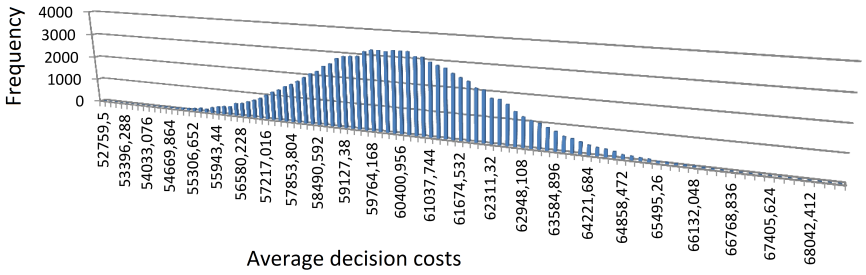


Figure 1 Average values of decision costs of the variant 1

Variant	Average	Expert 1	Expert 2
1	60 144	57 822	60 191
2	58 426	57 817	60 201
3	41 384	37 587	45 233
4	41 395	37 574	45 214

Table 1 Mean values of decision costs by variants of calculation [CZK]

Table 1 shows resultant mean values of decision costs by variants described in the previous paragraph. [7] contains more results of the example. Each calculated costs can be interpreted graphically as it is depicted in Figure 1. For each calculated costs, another statistical characteristics can also be counted up, e.g. minimum, maximum or variance. The characteristics can be helpful to support decision-making about establishment of the spare-parts stock. Some interesting properties of results are described in the next section. After that, some ideas of decision rule creating are stated.

2.1 Results properties based on the variants

Results obtained for each variants are not identical as can be easily seen from Table 1. Some of them are similar, which was proven by Welch test in R at a significance level $\alpha = 0.05$. The assumption of the test, e.g. normal distribution of the samples, let us assume fulfilled as it shows in Figure 1 whose decision costs have approximately normal distribution. Each decision costs have very similar distribution.

Resultant p-values of tests of the mean values equality for the first expert are stated in Table 2. The results of the second expert was very similar as for the first one. Statistically different results are only for the different rounding. Approach to the exact triangular distribution parameters are not included in the calculation. That is why, it should not have influence results of every expert which was affirmed by the tests. So, variants $1-2$ and $3-4$ are statistically undistinguishable.

Table 3 shows resultant p-values of tests of the mean values equality averaged for both experts. In this case, statistically different results depend on the different ways of rounding, and for different approaches to exact estimates of the triangular distribution parameters, results differ only for the rounding up. Therefore, only variants $3-4$ are statistically undistinguishable.

Variant	1	2	3	4
1	X	0.5531	< 2.2e-16	<2.2e-16
2	0.5531	X	< 2.2e-16	2.2e-16
3	< 2.2e-16	< 2.2e-16	X	< 0.4867
4	< 2.2e-16	< 2.2e-16	0.4867	X

Table 2 Testing p-values of the mean values equality of decision costs for the first expert

Variant	1	2	3	4
1	X	< 2.2e-16	< 2.2e-16	<2.2e-16
2	< 2.2e-16	X	< 2.2e-16	< 2.2e-16
3	< 2.2e-16	< 2.2e-16	X	0.5859
4	< 2.2e-16	< 2.2e-16	0.5859	X

Table 3 Testing p-values of the mean values equality of decision costs for the average of both experts

2.2 Decision-making rules

The right decision should be based on good information. Data can be obtained from the model. However, knowledge, which is necessary to add to data to achieve the information, should be gained from managers and experts. Nevertheless, sometimes it is good to get some advices to rules construction from the model creator. In this section, two ideas are described: based on the net present value (NPV) and coming out from financial statements.

The first decision rule uses idea of NPV which compare investment costs and expected incomes from the investment. Both is discounted to the present time. However, company, which provided data used in the example, did not have suitable data in their records. Therefore, I have modified this approach. It is possible to compare calculated decision costs for a year with an average yearly fine for the repairs delay. A value of the fine should be taken from the records of the company. Nevertheless, the company did not have registered it. So, another approach was needful.

In the second way, *financial statements* can be used as base to the decision. A company balanced-sheet is important for it. If there are excessive reserves in the balance-sheet, then the company should have

enough money to establish the spare-parts stock. However, the company from the example did not have any excessive reserves. So, I do not have any information if the company have free resources, which can be invested into spare-parts.

The decision about the spare-parts stock is still at the company management. *Relationship of the decision-maker to the risk* can be considered to improve described decision rules. Nevertheless, if the decision is not possible using the rules, then, decision-maker relationship to the risk is unimportant.

3 Possibilities of a further research

The model described in the paper can be improved. The first idea is to take *shorter time period* into account. Then the formulas should come out from well-known classical inventory models which are described e.g. in [3] and in [4].

The second possibility is to use *fuzzy sets* instead of the triangular distribution. In this case, the demand calculation should use interval arithmetic or extension principle. The third idea is using *fuzzy Monte Carlo*. It is Monte Carlo simulation where parameters of statistical distributions are fuzzy numbers. The largest problem of this approach could be obtaining fuzzy parameters estimations from experts. The calculation will be more difficult than for the classical Monte Carlo simulation.

However, changes of the model do not make sense because its validation might be impossible to do. The main problem is that the time period of the equipment complex field test have not finished, yet. So, I do not have suitable data to validate the model. I hope that I will obtain them in the next year.

4 Conclusion

The model was introduced and described. Usage of the model was demonstrated in the example. The example was complemented by tests of the mean values equality and explaining some possible decision rules. Some further research possibilities were stated.

A validation of gained results is necessary for being able to finish the model. After the validation, the best calculation variant should be chosen based on the validation results. If the variant, which uses the modification of exact parameters estimations, was selected, then, a sensitivity analysis of the modifying constant influence should be done.

Acknowledgements

The paper is supported by the grant IGA PrF 2014028 Mathematical Models of the Internal Grant Agency of the Palacký University in Olomouc.

References

[1] Gentle, J. E.: *Statistics and Computing: Random Number Generation and Monte Carlo Methods*. 2nd edition. Springer, New York, 2003.

[2] Gros, I.: *Kvantitativní metody v manažerském rozhodování*. Grada, Praha, 2003.

[3] Hillier, F. S., Liebermann, G. J.: *Introduction to operations research*. 7th edition. McGraw-Hill, New York, 2001.

[4] Hušek, R., Maňas, M.: *Matematické modely v ekonomii*. SNTL, Praha, 1989.

[5] Kenedy, W. J., Patterson, J. W., Fredendall, L. D.: An overview of recent literature on spare parts inventories, *International Journal of Production Economics* **2** (2002), 201 – 2015.

[6] Kuncová, M.: Možnosti využití kvantitativních metod a simulací při řízení zásob v dodavatelských řetězcích1, *Statistika* **4** (2006), 317 -– 326.

[7] Melicheríková, P.: *Aplikace matematických modelů ve strojírenském podniku se zaměřením na tepelné elektrárny a skládkové hospodářství*. Palacký University Olomouc, Olomouc, 2013.

Application of an Asymmetric Banzhaf Power Index: The Case of the Czech Parliament

Elena Mielcová[1]

Abstract. In general, parliamentary voting can be described as a cooperative game with transferable utility function with political parties as agents of the game. Studies of decision-making process of parliamentary voting from the mathematical point of view usually cover studies on power and influence of coalitions. As a measure of players' power, the concept of power indices was introduced in the second half of the last century. One of the widely used power indices – Banzhaf power index – also called Penrose-Banzhaf or Banzhaf-Coleman power index – is based on the concept of 'swing voters' in possible coalitions. The Banzhaf power index is derived from the assumption of random voting. However, real world results differ from theoretical assumptions. For example Gelman, Katz and Bafumi (2004) and Gelman, Katz and Tuerlinckx (2002) have shown that the assumption is violated in large voting bodies. In order to fit real world conditions, adjustments of the index were done by Hu (2006), who proposed add weighted marginal contributions to power indices. This article applies Hu's ideas to calculate power in a real voting, namely the data from the Chamber of Deputies of the Czech Parliament with the emphasis on the State Budget voting issues.

Keywords: Banzhaf power index, asymmetric power indices, parliamentary voting, coalitions.

JEL Classification: C71
AMS Classification: 91B12

1 Introduction

Decision-making process of parliamentary voting can be modelled as a simple cooperative game with transferable utility function. Studies of cooperative games also cover idea of evaluation of players of such a game. The first attempt to evaluate power of players in voting game was done by Penrose (1946) [8], who tried to express the power of a player by a probability to be in a winning coalition under the condition that other players vote randomly. Later on, Banzhaf (1965) [1], used Penrose idea of voters' power to discuss fairness of Nassau county board voting system. He described the power as the ratio of so-called swing voters of winning coalitions, and he demonstrated the unfairness of such a voting system to county inhabitants. This power index was applied by J. S. Coleman (1971) [3], who added idea of normalization to power measure calculations. Now, the Banzhaf power index is often called Penrose-Banzhaf or Banzhaf-Coleman power index. Meanwhile, the transformation of now well-known Shapley value [9] to the context of simple cooperative games was done by Shapley and Shubik in 1954 [10]. The normalized Banzhaf power index together with the so-called Shapley-Shubik power index till now serves as a-priori evaluation of coalitions in voting bodies.

In real world, agents of the simple cooperative game – usually political parties – act not strictly as predicted in theory. That means, calculated a-priori power indices do not reflect actual distribution of power in voting bodies. As all real systems are full of an uncertainty, also parliamentary voting can be described up to some degree of freedom. For example Gelman, Katz and Bafumi (2004) [5] have shown that the assumption of randomly distributed votes is violated in large voting bodies; their observations are based on data from the Electoral College of the United States, as well as other voting bodies within the U.S. and Europe. Gelman, Katz and Tuerlinckx (2002) [4] tried to explain difficulties in choosing the right voting rule for a fair voting power distribution in U.S. Presidential votes by states. In order to fit real world conditions, various adjustments of the Banzhaf index were done; for example Hu (2006) [7] proposed add weighted marginal contributions to power indices. The main aim of this article is to test appropriateness of Hu's adjustments to the normalized Banzhaf index on the real world data from the Czech Parliament. Thus, this article applies the concept of additional weighted marginal contributions to normalized Banzhaf power index in order to calculate power in a real voting,

[1] Silesian University in Opava / School of Business Administration in Karvina, Department of Mathematical Methods in Economics and CERGE-EI Affiliate Fellow, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech Republic, mielcova@opf.slu.cz.

namely the data from the Chamber of Deputies (Lower House) of the Czech Parliament with the emphasis on the State Budget voting issues. This article is structured as follows: The next part is devoted to description of theoretical issues of both a-priori normalized Banzhaf power index and a concept of asymmetric Banzhaf power index, followed by the part describing data and calculated results. Conclusions and list of references ends the text.

2 Banzhaf Power Index

Any parliamentary voting is usually described as a simple monotonic game with transferable utility function. Formally, if U denotes the universe of players then $N \subset U$ is a set of players in voting game; a transferable utility game $\Gamma(N)$ on N is a mapping of all subsets (in voting also called coalitions) of N, we denote $v: 2^N \to \mathbf{R}$, such that $v(\varnothing) = 0$. Function $v(x)$ is called a characteristic function of a game. A transferable utility game v on N is simple, if the characteristic function can obtain only values 0 or 1: $\forall S \subset N; v(S) = 0 \vee v(S) = 1$. A simple game on N is monotonic if $v(S) \leq v(T)$ for all $S,T \subset N$ such that $S \subset T$.

Let W denotes the set of winning coalitions. According to [7], the simple game, denoted by $\Gamma(N,W)$, with characteristic function $v: 2^N \to \{0,1\}$ is the game that satisfies three conditions: (*i*) the empty coalition never wins: $v(\varnothing) = 0$, (*ii*) the grand coalition always wins: $v(N) = 1$, and (*iii*) any superset of winning coalition also wins. The characteristic function of such a game is defined as $v(T) = 1$ if $T \in W$ and $v(T) = 0$ otherwise.

The idea behind the Banzhaf index is based on calculations of players that are crucial for a coalition to be winning - so called swing players. Let player i be in a coalition T. Denote $T \setminus i$ be a coalition that is created from T by omission of player i. For any $T \subset N$ we say player i swings in T if $T \in W$, and $(T \setminus i) \notin W$. If i is a swing player of a coalition T in a simple monotonic game, then $v(T) - v(T \setminus i) = 1$. The originally proposed Banzhaf index [1, 3] counts the number of swings:

$$b_i = \frac{1}{2^{|N|-1}} \sum_{T \subseteq N,\, i \in T} \left(v(T) - v(T \setminus i) \right) \tag{1}$$

where the cardinality of the set N is denoted by $|N|$. According to Haller (1994) [6], the Banzhaf value is the unique value function on the monotone transferable utility games that satisfies linearity, the dummy player property, anonymity, and the proxy agreement property.

The standardized Banzhaf index, also called normalized Banzhaf index [1, 3] calculates voter i's Banzhaf index to sum of all Banzhaf indices for all players. The normalized Banzhaf index value for voter i is obtained by dividing the sum of i's swings (regarding all possible combinations) by the sum of all voters' all swings. Formally voter i's standardized Banzhaf index is calculated as

$$b_i^{|N|} = \frac{b_i}{\sum_{j \in N} b_j} = \frac{\frac{1}{2^{|N|-1}} \sum_{T \subseteq N,\, i \in T} \left(v(T) - v(T \setminus i) \right)}{\sum_{j \in N} \frac{1}{2^{|N|-1}} \sum_{T \subseteq N,\, j \in T} \left(v(T) - v(T \setminus j) \right)} = \frac{\theta_i}{\theta} \tag{2}$$

where θ_i denotes the number of voter i's swings, θ is a total number of swings of all players. The normalized Banzhaf value satisfies anonymity, however, it does not satisfy linearity, the dummy player property, and the proxy agreement property [2].

In the definition of Banzhaf index there are weighted marginal contributions $v(T) - v(T \setminus i)$ in forming winning coalitions from losing ones. Hu (2006) [7] proposed to take into account also blocking coalitions' power. Let a simple game, denoted by $\Gamma^*(N,W^*)$ is a game of such a blocking coalition – a winning coalition T in Γ^* blocks forming any winning coalition in Γ. Because $W^* = \{T; N \setminus T \notin W\}$, the $\Gamma^*(N,W^*)$ can be considered a dual of $\Gamma(N,W)$ with characteristic function $v^*(T) = 1 - v(N \setminus T)$. Let $\mathbf{S}$, and $\mathbf{S}^*$ denote the coalition of players who vote for , and against a bill, respectively. The asymmetric Banzhaf index is based on so-called double swing players, which means the players of T that are swing players in either $\Gamma(N,W)$ or $\Gamma^*(N,W^*)$ [7]:

$$\bar{b}_i = \sum_{T \subseteq N,\, i \in T} P_T \left(v(T) - v(T \setminus i)\right) + \sum_{T \subseteq N,\, i \in T} P_T^* \left(v^*(T) - v^*(T \setminus i)\right) \tag{3}$$

In the last equation, $P_T = \text{Prob}(\mathbf{S} = T)$ and $P_T^* = \text{Prob}(\mathbf{S}^* = T)$. When $\mathbf{S}$ is uniformly distributed on 2^N, then asymmetric Banzhaf index is the same as Banzhaf index [7].

In this paper, the calculated a-priori Banzhaf index, and normalized Banzhaf index is compared to asymmetric Banzhaf index and its normalization, where values of P_T, and P_T^* are based on relative frequencies of coalition formation of year-to-year state budget voting in the Chamber of Deputies of the Czech Parliament 2006-2013. All calculated power indices are compared to index of success of political parties in voting. In general, the political party success index is defined as the ratio of decisions of the Lower House that were the same as the party decisions to all decisions during the parliamentary period:

$$I_{success} = \frac{\text{number of political party decisions identical with parliamentary decisions}}{\text{number of parliamentary decisions}} \tag{4}$$

The party decision is derived from the votes of party members using simple majority rule. The coefficient of party success can reach the values from the interval $\langle 0,1 \rangle$.

3 Data Description and Results of Calculations

This study is based on the voting on state budget issues in the Chamber of Deputies of the Czech Parliament during two parliamentary periods. In the Czech Parliament, the outcome of every vote for every member can be "no", "yes", "present, abstain", "absent". Every bill to be passed needs at least as many "yes" votes as quota. As the quota is based on the sum of all present legislators, the outcome "present, abstain" in fact serves as "no" outcome; hence, in this analysis, this outcome is reclassified to "no" outcome. Basic information on the Czech Parliamentary system as well as the set of all historical votes can be found at the official web site of the Lower House of the Czech Parliament URL: www.psp.cz [12].

During the 2006-2010 parliamentary period, there were five political parties operating in the Lower House of the Czech Parliament; three political parties of the governmental coalition: Civic Democratic Party (ODS), Christian and Democratic Union – Czechoslovak People's Party (KDU-ČSL), and Green Party (SZ), and two political parties, which stayed in opposition: Czech Social Democratic Party (ČSSD) and Communist Party of Bohemia and Moravia (KSČM). 2006-2010 Parliament was elected in June 2006, however the government was approved after months of discussions in January 2007. After the government was set, the governmental coalition composed of ODS, KDU-ČSL, and SZ worked together. Differences in political programs and disagreement between governmental political parties ended up in a downfall of the government; on March 24, 2009, the Lower House of the Czech Parliament approved no-confidence of the government. The caretaker government was set and functioned till elections in 2010. Detailed description of the situation after 2006 Parliamentary Elections as well as of the non-confidence voting can be found in [11].

During the 2010-2013 parliamentary period, there were five political parties operating in the Czech Parliament: ODS, TOP-09 and Věci veřejné (VV) as governmental coalitional parties together with ČSSD and KSČM, which stayed in opposition. The Government was set in July 2010. In April 1011, two deputies, members of VV were accused of briberies. This revelation started parliamentary crises that peaked in leaving more than half of Lower House deputies from VV; moreover VV became an oppositional party. The Government was forced to resign on 17 June 2013, after another corruption and bribery scandal. A newly set caretaker government resigned in August 2013. The Chamber of Deputies passed a motion dissolving itself on August 20; the elections were scheduled for 25 and 26 October 2013 [13].

For illustration, Tables 1 and 2 give values of a-priori Banzhaf indices, and normalized Banzhaf indices as well as values of party success indices for budget voting for political parties present in 2006-2010, and 2010-2013 parliaments, respectively. In both tables, political parties are sorted with respect to number of seats. The interesting point is a fact that the coefficient of party success is not always bound with a-priori Banzhaf index; for rough comparison, the correlation coefficient of an a-priori Banzhaf index (as well as a-priori normalized Banzhaf index) with party success index is -0.28 and -0.06 for 2006-2010 and 2010-2013 parliaments, respectively, both correlation coefficients are not statistically significant at 5% level of significance. Even when comparing ranks of calculated Banzhaf indices with ranks of party success coefficients, the calculated Spearman's rank correlation coefficient gives statistically not significant results (r=0 and r=-0.06 for the first and the second Chamber of Deputies, respectively). However, taking into account a-priori coalitional structure, a coefficient of party success

is correlated with an a-priori Banzhaf index with given coalitional structure; the correlation coefficient between the two variables is 0.98, this correlation coefficient is statistically significant at 5% level of significance. Hence, we could expect strong party discipline of governmental coalition parties in 2010-2013 state budget voting.

	ODS	ČSSD	KSČM	KDU-ČSL	SZ
Seats	81	74	26	13	6
A-priori Banzhaf index	0.28125	0.21875	0.21875	0.03125	0.03125
A-priori normalized Banzhaf index	0.36	0.28	0.28	0.04	0.04
Party success index	0.7411	0.9045	0.4607	0.8829	0.7273

Table 1 Number of seats, a-priori Banzhaf index and normalized Banzhaf index for political parties after 2006 Chamber of Deputies Elections; index of party success for state budget voting during 2006-2010 parliamentary period. Source: own calculations.

	ČSSD	ODS	TOP-09	KSČM	VV
Seats	56	53	41	26	24
A-priori Banzhaf index	0.25	0.25	0.125	0.125	0.125
A-priori normalized Banzhaf index	0.2857	0.2857	0.1429	0.1429	0.1429
A-priori normalized Banzhaf index with coalition structure	0	0.33	0.33	0	0.33
Party success index	0.2468	0.9740	0.9740	0.2078	0.7792

Table 2 Number of seats, a-priori Banzhaf index and normalized Banzhaf index for political parties after 2010 Chamber of Deputies Elections; index of party success for state budget voting during 2010-2013 parliamentary period. Source: own calculations.

Calculations of spatial Banzhaf power index are based on all possible winning coalitions as well as all possible blocking coalitions. The interesting point is the fact, that in the beginning of 2006-2010 parliamentary period, there were fifteen winning coalitions, but seventeen blocking coalitions – the consequence of difficult situation at the beginning of electoral period, when coalitional political parties – ODS, KDU-CSL, and SZ gained 100 seats in Parliament, and oppositional political parties – KSCM and KSCM had gained 100 seats, as well. In the Chamber of Deputies, when all representatives are present, the quota is 101 votes, thus the situation after 2006 Elections led to deadlock. Summation of all winning and blocking coalitions together with swing players for 2006-2010 parliamentary period is given in Table 3.

Winning or blocking coalition	Win	Block	Swing players winning	Swing players blocking	Winning or blocking coalition	Win	Block	Swing players winning	Swing players blocking
A+B	Yes	Yes	A, B	A, B	B+C+D	Yes	Yes	B,C,D	B, C
A+C	Yes	Yes	A, C	A, C	B+C+E	Yes	Yes	B, C, E	B, C
B+C	No	Yes	-	B, C	A+B+C+D	Yes	Yes	-	-
A+B+C	Yes	Yes	A	-	A+B+C+E	Yes	Yes	-	-
A+B+D	Yes	Yes	A, B	A, B	A+B+D+E	Yes	Yes	A, B	A
A+B+E	Yes	Yes	A, B	A, B	A+C+D+E	Yes	Yes	A, C	A
A+C+D	Yes	Yes	A, C	A, C	B+C+D+E	Yes	Yes	B, C	B, C
A+C+E	Yes	Yes	A, C	A, C	A+B+C+D+E	Yes	Yes	-	-
A+ D+E	No	Yes	-	A,D,E					

Table 3 List of winning and blocking coalitions and swing players in 2006-2010 Chamber of Deputies; A=ODS, B=CSSD, C=KSCM, D=KDU-CSL,E=SZ. Source: Own calculations.

The situation at the beginning of 2010-2013 parliamentary period was not so difficult, seats to five parliamentary political parties were distributed such that all winning coalitions were also the blocking ones, and even all swing players of winning coalitions were swing players of the blocking coalitions. Summation of all winning and blocking coalitions together with swing players for 2006-2010 parliamentary period is given in Table 4.

Winning or blocking coalition	Swing players	Winning or blocking coalition	Swing players
A+B	A, B	B+C+E	B, C, E
A+B+C	A, B	B+D+E	B, D, E
A+B+D	A, B	A+B+C+D	-
A+B+E	A, B	A+B+C+E	-
A+C+D	A, C, D	A+B+D+E	-
A+C+E	A, C, E	A+C+D+E	A
A+ D+E	A, D, E	B+C+D+E	B
B+C+D	B, C, D	A+B+C+D+E	-

Table 4 List of winning and blocking coalitions and swing players in 2010-2013 Chamber of Deputies; A=CSSD, B=ODS, C=TOP-09, D=KSCM, E=VV. Source: Own calculations.

Calculations of spatial Banzhaf power index are based on all possible winning and blocking coalitions, together with counting relative frequencies for all winning and blocking coalition. The coalitional structure was expected to be the same as at the beginning of the respective parliamentary period. As proven in [7], if true probabilities over such a parliament were distributed uniformly over all coalitions, the calculated index is the same as an a-prori Banzhaf index as given in Table 2. However, this is usually not valid in the real voting body; for illustration, Figure 1 gives a comparison of uniformly distributed votes with real relative frequencies in state budget voting through overall 2006-2010 period.

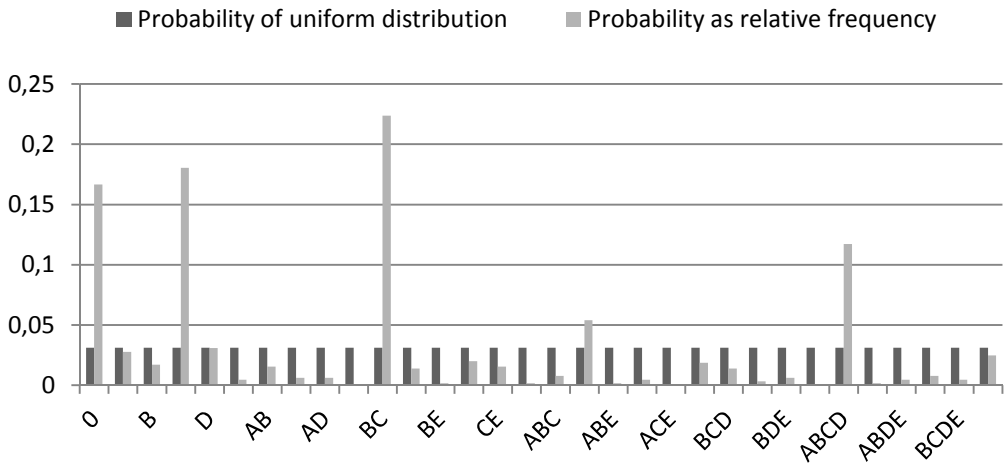


Figure 1 Comparison of uniformly distributed votes with real relative frequencies in state budget voting in 2006-2010 period; A=ODS, B=CSSD, C=KSCM, D=KDU-CSL, E=SZ. Source: Own calculations.

	ODS	CSSD	KSCM	KDU-CSL	SZ
2007 State Budget	0.5417	0.2667	0.1111	0.0472	0.0333
2008 State Budget	0.3636	0.1000	0.1091	0.2090	0.2182
2009 State Budget	0.3434	0.0151	0.0151	0.3133	0.3133
2010 State Budget	0.3594	0.1719	0.1875	0.1719	0.1094
	CSSD	ODS	TOP-09	KSCM	VV
2011 State Budget	0.0123	0.3210	0.3210	0.0123	0.3333
2012 State Budget	0	0.3469	0.3265	0	0.3265
2013 State Budget	0.2889	0.0444	0.0444	0.2889	0.3333

Table 5 Calculated asymmetric normalized Banzhaf power of political parties during 207-2013 state budget voting. Source: Own calculations.

The situation in the Chamber of Deputies of the Czech Parliament is reflected on calculated power of political parties in state budget voting 2007-2013 (Table 5). In 2007 state budget voting, the cooperation of opposi-

tional political parties ODS and CSSD almost ruled out other coalitional parties – KDU-CSL and SZ. In the 2008 and 2009 state budget voting, the influence of CSSD decreased, influence of KDU-CSL and SZ substantially increased. The decline in voting discipline of governmental parties in 2010 state budget voting caused decrease in power of all three governmental parties and increase of influence of oppositional political parties – CSSD and KSCM. Strong party discipline of governmental coalitional political parties in 2010-2013 parliamentary period gave substantial part of voting power to coalitional political parties – this power is almost the same as an a-priori Banzhaf index with known coalitional structure. Calculated power of political parties in 2013 state budget voting is to some extent false – VV lost its power due to decrease in number of seats, however values of characteristic function are calculated with respect to original values, hence the calculated indices for VV, CSSD and KSCM are higher than in reality. The correlational coefficient of an average normalized asymmetric Banzhaf power index with index of success is 0.178 and 0.812 for the 2006-2010 and 2010-2013 parliamentary periods, respectively; the second one is statistically significant on 5% level of significance, there is an increase in both calculated correlation coefficients.

4 Conclusion

This article compares the calculated a-priori normalized Banzhaf power index with asymmetric normalized Banzhaf power index of political parties based on relative frequencies of coalitions for budget voting data from the 2006-2013 electoral period. The calculated results of power distribution show differences through years, moreover, these indices are better comparing to a-priori Banzhaf indices: correlational coefficients of average normalized asymmetric Banzhaf power index with index of success are higher than correlation indices of an a-priori normalized Banzhaf power index with an index of success.

Acknowledgements

This research was supported by the GACR project No. 14-02424S.

References

[1] Banzhaf, J. F.: Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, **19** (1965), 317–343.

[2] Brink, J.R. van den, Laan, G. van der: The Normalized Banzhaf Value and the Banzhaf Share Function. *International Journal of Mathematics, Game Theory and Algebra*, **9**(1999), 65-84.

[3] Coleman, J. S.: Control of Collectives and the Power of a Collectivity to Act. In: Social Choice (Lieberman B., ed.), Gordon and Breach, New York, 1971, 192–225.

[4] Gelman, A., Katz, J. N., Tuerlinckx, F.: The Mathematics and Statistics of Voting Power. *Statistical Science: Voting and Elections*, **17** (2002), 420-435.

[5] Gelman, A., Katz, J. N., Bafumi J.: Standard Voting Power Indexes Do Not Work: An Empirical Analysis. *British Journal of Political Science*, **34** (2004), 657-674.

[6] Haller, H.: Collusion Properties of Values. *International Journal of Game Theory*, **23**, (1994), 261-281.

[7] Hu, X.: An asymmetric Shapley-Shubik power index. *International Journal of Game Theory*, **34** (2006), 229–240.

[8] Penrose, L. S. The Elementary Statistics of Majority Voting. *Journal of the Royal Statistical Society*, **109** (1946), 53-57.

[9] Shapley L. S.: A value for n-Person Games. In: *Contribution to the Theory of Games* (Kuhn H., and Tucker A. W., eds.) vol. 2, Princeton University Press, 1953, 307-317.

[10] Shapley L. S. and Shubik M.: A Method for Evaluating the Distribution of Power in a Committee System. *The American Political Science Review* **48** (1954), 784-792.

[11] Škochová, Marie. Analýza voleb do Poslanecké sněmovny Parlamentu České republiky v roce 2006 [online]. *E-polis.cz*, 15. srpen 2008. [cit. 2014-05-05]. Available at WWW: <http://www.e-polis.cz/komparace/297-analyza-voleb-do-poslanecke-snemovny-parlamentu-ceske-republiky-v-roce-2006.html>. ISSN 1801-1438.

[12] Web: *Poslanecká sněmovna Parlamentu České republiky* [online]. [cit. 2014-05-05]. Available at WWW: http://www.psp.cz/sqw/hp.sqw

[13] Wikipedia: Seznam hlasování o důvěře vládě České republiky. [online]. [cit. 2014-05-05]. Available at WWW:<http://cs.wikipedia.org/wiki/Seznam_hlasov%C3%A1n%C3%AD_o_d%C5%AFv%C4%9B%C5%99e_vl%C3%A1d%C4%9B_%C4%8Cesk%C3%A9_republiky#Poslaneck.C3.A1_sn.C4.9Bmovna_P.C4.8CR.2C_5._volebn.C3.AD_obdob.C3.AD_2006.E2.80.932010>.

Minimum wage setting mechanism and its effects

Stepan Mikula[1], Jaroslav Schulz[2]

Abstract. There is a great number of empirical papers which deals with the effect of minimum wage on unemployment. However those studies usually do not take into account the way of minimum wage setting which may be the source of its heterogeneous effects. Using panel data from OECD Statistical database (period of 1985-2011) we investigate relationship between unemployment and minimum wage in developed countries. In our analysis we focus on different effects of differently set minimum wages. We show that the minimum wage setting mechanism matters with respect to the effect of minimum wage on unemployment.

Keywords: Minimum wage, unemployment, setting mechanism.

JEL classification: J64
AMS classification: 91B40

1 Introduction

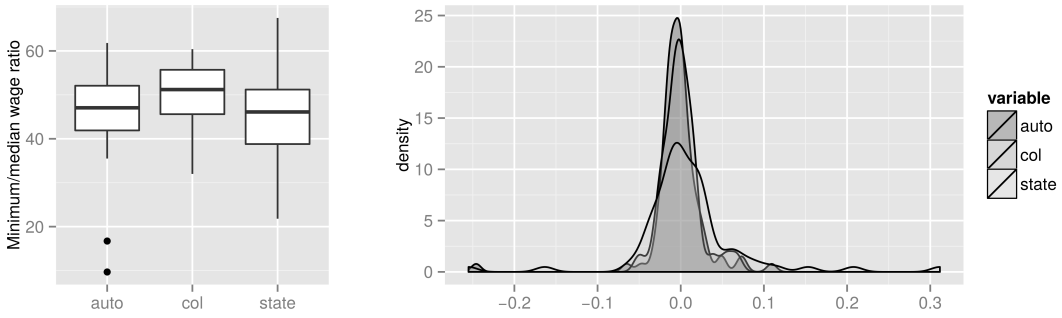
Research on minimum wage and its impacts on various tasks has been very intensive for last 70 years. The most prominent topic was the effect of minimum wage on unemployment. The conclusions of this research underwent tumultuous evolution with an ambiguous result. Well-arranged overview presents [20]. Theoretical as well as empirical literature dominantly describes the minimum wage as a wage floor based on laws and regulations. However, wages resulting from bargaining between unions and firms are also considered to be minimum wages and they are incorporated in the research works without extra.

We are convinced that labor market analysis belongs to one of the few areas in economics which are very deeply influenced by laws and regulations. Without any concerns we could pronounce, that each country has its own specific labor legislation that creates different institutions and lead to different outcomes on labor markets and hence economic performance. Few papers from [23], [21], [9], [14], [5] or [10] were dedicated to various impacts of minimum wage legislation. All of them, however, associate minimum wage laws just with a binding minimum wage rate and then derive impacts of such binding minimum wage on diverse variables.

Nevertheless, the process of minimum wage setting might be one of the important factors when deriving the impacts of minimum wage on unemployment. There is scarce literature in economics about the process leading to determination of a particular minimum wage level. As a pioneer has to be considered [4], who examines that fixing regime of minimum wage matters. He introduces a simple theoretical model and empirical calibration. He classifies three minimum wage setting regimes – collective bargaining, unilaterally by government and government decree consulted with third parties. However the aim of his article is the relation between minimum wage level and wage setting regime. He finds out that a government legislated minimum wage is consistently lower than a wage floor set within collective agreement.

Unlike Boeri, we investigate the impact of minimum wage including wage setting regime on unemployment. We also divide the formation process into three categories. In fact, it does not matter whether the government consults the decision of minimum wage variation or not. It is still unilateral act in law executed by government decree or law. But what matters are auto adjustment institutions in minimum wage law, such as indexation. Therefore we find more appropriate to identify three minimum wage setting regimes different to Boeri – collective bargaining, government unilateral decision with auto adjustment elements and government unilateral decision without any auto adaptation.

[1]Masaryk University, Department of economics, Lipova 41a, Brno, Czech Republic, mikula@econ.muni.cz
[2]Masaryk University, Department of economics, Lipova 41a, Brno, Czech Republic, 150971@mail.muni.cz



(a) Minimum wage distribution

(b) Density of the minimum wage growth rate

Figure 1: Minimum wage

Our data (see Section 2 and Figure 1a) are consistent with the findings of Boeri. One-tailed Mann-Whitney U test rejects the null in favor of alternative hypothesis that minimum wages set within collective bargaining are greater then government-set minimum wages ($W = 5869.5^{***}$ for government-set minimum wage with auto-adjustment elements and $W = 5506.5^{***}$ without).

Although according to International Labor Organization [3] provides the minimum wage by law 90 % of world countries, there are important countries where the minimum wages are generally set by various collective bargaining.

Collective wage agreements determined by collective bargaining are common in developed countries with advanced social dialogue and mostly high labor union density. Although there are some theoretical papers on collective bargaining [11], it is complicated to model such situation. Minimum wages set by collective bargaining seem to be in some ways an endogenous issue, which on the other hand does not correspond to reality, where it is more psychological issue from the game theory field. Based on fundamental work from [6], we distinguish centrally organized or coordinated collective bargaining and the opposite. Centrally organized bargaining leads to lower wage level structure in comparison to decentralized. The intuition behind is that centrally organized unions take macroeconomic performance and indicators into account which yields in productivity-oriented claims.

On the other hand there is a vast theoretical literature on minimum wage impacts on various factors. The conclusions are ambiguous and seem to be conforming to chosen theoretical framework. Albeit it seems that the minimum wage set by government does not have under certain thresholds significant effect on economy, wage distribution, poverty etc. the opposite could happen via spillover effects based on wage contour theory [15].

From above mentioned, it is clear, that the labor economics is affected by a huge variety of factors that cannot be caught by theory. Therefore we share the opinion of [18] that the questions about economic effects of minimum wage need to be resolved empirically. Economic theory is a valuable tool in suggesting testable hypothesis as well as potential impacts.

2 Model and data

Our empirical model is designed to asses the influence of minimum wage on unemployment. More precisely we investigate the relationship between differences of unemployment rate and growth rate of minimum wage. We use following base line model specification:

$$\Delta u_{i,j} = \beta_{\mathbf{0}} + \beta_1 g_{i,j} + \beta_2 g_{i,j-1} + \beta_3 inst_{i,j} + \beta_4 \Delta benefits_{i,j} + \varepsilon_{i,j} \tag{1}$$

The first differences in unemployment rate in country i and year j ($\Delta u_{i,j}$) are explained by the growth rate of GDP ($g_{i,j}$) and its first lag ($g_{i,j-1}$); strictness of labor market institutions ($inst_{i,j}$) and first difference in unemployment benefits measured as percent of GDP ($\Delta benefits_{i,j}$). We perform estimation of the pooled model and estimation with country and time fixed effects. (The composition of the vector of constants ($\beta_{\mathbf{0}}$) is therefore different for each model specification.)

We use the yearly data from OECD database for 20 European countries (for list see below) in period of 1985–2011.

The base-line model (1) is extended in the following way to pick up the effect of the minimum wage growth:

$$\Delta u_{i,j} = \beta_{\mathbf{0}} + \beta_1 g_{i,j} + \beta_2 g_{i,j-1} + \beta_3 inst_{i,j} + \beta_4 \Delta benefits_{i,j} + \beta_5 wage_{i,j} + \varepsilon_{i,j} \quad (2)$$

$$\Delta u_{i,j} = \beta_{\mathbf{0}} + \beta_1 g_{i,j} + \beta_2 g_{i,j-1} + \beta_3 inst_{i,j} + \beta_4 \Delta benefits_{i,j} + \\ + \beta_6 col_i \times wage_{i,j} + \beta_7 auto_i \times wage_{i,j} + \beta_8 state_i \times wage_{i,j} + \varepsilon_{i,j} \quad (3)$$

The variable $wage_{i,j}$ depicts the change of minimum wage. We describe it as the growth rate of the share of minimum wage on the median wage. The variable *wage* is in equation (3) multiplied by dummy variables which reflects the mechanism of minimum wage setting:

- variable *col* is equal to one if the minimum wage is set by collective bargaining (it holds for Austria, Belgium, Estonia, Finland, Germany, Italy, Sweden and Greece)
- variable *auto* is equal to one if the minimum wage is set by the government in the presence of automatic adjustment mechanism (it holds for France, Luxembourg, Slovenia, Poland, Spain and the Netherlands)
- variable *state* is equal to one if the minimum wage is set by the government without any automatic adjustment mechanism (it holds for Czech Republic, Ireland, Portugal, United Kingdom, Hungary and Slovakia)

Countries are classified on the bases of ILO's TRAVAIL database, which compares single conditions of Work and Employment laws.[1] Countries categorized under *auto* variable incorporated auto-adjustment clause provisions in their Labor Legal Code. The adjustment is implemented via mainly two variables indexation – change of CPI or GDP growth rate.

Summary statistics of all variables used are given in Table 1.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Δu	0.048	1.456	−4.358	−0.654	−0.092	0.575	8.217
g	2.813	3.564	−10.293	0.946	3.144	4.912	13.351
$inst$	2.632	0.938	1.198	2.004	2.357	3.027	5.000
$\Delta benefits$	0.013	0.137	−0.300	−0.050	0.000	0.068	0.460
$wage$	0.002	0.048	−0.255	−0.013	−0.002	0.013	0.312
$col \times wage$	−0.000	0.009	−0.050	0.000	0.000	0.000	0.067
$auto \times wage$	−0.001	0.023	−0.246	0.000	0.000	0.000	0.109
$state \times wage$	0.003	0.041	−0.255	0.000	0.000	0.000	0.312
col	0.182	0.387	0	0	0	0	1
$auto$	0.433	0.497	0	0	0	1	1
$state$	0.384	0.488	0	0	0	1	1

Table 1: Summary statistics

3 Results

The estimated parameters are presented in Table 2.[2] The estimated results indicate the presence of robust significant effect of minimum wage changes in the case of regulations set by government. However the effects are opposite. The decrease of the distance between minimum and median wage increases the unemployment in the presence of automatic adjustment mechanism and has an opposite effect in case of its absence.

The literature provides possible explanations for both effects. The non-negative impact of minimum wage variation on unemployment emphasize particularly [16], [13],[22] or [7]. Such results are explained

[1] http://www.ilo.org/travail/lang–en/index.htm

[2] The `R` and `plm` package [19, 8] are used.

	Dependent variable: First difference of unemployment rate ($\Delta u_{i,j}$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
g_j	-0.207^{***}	-0.209^{***}	-0.203^{***}	-0.201^{***}	-0.203^{***}	-0.208^{***}	-0.128^{***}	-0.127^{***}	-0.133^{***}
	(0.039)	(0.039)	(0.023)	(0.034)	(0.034)	(0.035)	(0.041)	(0.041)	(0.041)
g_{j-1}	-0.086^{***}	-0.085^{***}	-0.088	-0.120^{***}	-0.120^{***}	-0.113^{***}	-0.159^{***}	-0.162^{***}	-0.151^{***}
	(0.020)	(0.021)	(0.223)	(0.022)	(0.023)	(0.023)	(0.030)	(0.033)	(0.032)
$inst$	0.140^{**}	0.152^{**}	0.130	0.611^{**}	0.635^{***}	0.572^{**}	0.428^{**}	0.433^{**}	0.407^{**}
	(0.055)	(0.059)	(1.124)	(0.252)	(0.239)	(0.223)	(0.192)	(0.196)	(0.185)
$\Delta benefits$	1.945^{**}	1.998^{**}	1.937	1.312	1.343	1.374	3.004^{***}	3.035^{***}	2.971^{***}
	(0.789)	(0.830)	(7.976)	(1.027)	(1.112)	(1.124)	(0.971)	(1.041)	(1.001)
$wage$		0.957			0.373			-0.678	
		(1.979)			(1.856)			(1.436)	
$col \times wage$			20.021^{***}			-0.178			-0.750
			(3.227)			(7.976)			(7.056)
$auto \times wage$			-7.041^{***}			-6.695^{**}			-5.790^{*}
			(1.295)			(3.227)			(3.099)
$state \times wage$			2.568^{***}			2.911^{**}			1.189
			(0.035)			(1.295)			(1.146)
Constant	0.530^{***}	0.492^{***}	0.530^{***}						
	(0.158)	(0.158)	(0.035)						
Observations	206	204	204	206	204	204	206	204	204
R^2	0.456	0.457	0.490	0.463	0.459	0.477	0.622	0.617	0.626
Adjusted R^2	0.445	0.443	0.471	0.420	0.414	0.426	0.492	0.484	0.485
F Statistic	42.110^{***}	33.280^{***}	26.874^{***}	40.253^{***}	31.197^{***}	23.760^{***}	9.567^{***}	8.902^{***}	8.530^{***}
	(df = 4; 201)	(df = 5; 198)	(df = 7; 196)	(df = 4; 187)	(df = 5; 184)	(df = 7; 182)	(df = 28; 163)	(df = 29; 160)	(df = 31; 158)
Country fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	No	No	No	No	No	No	Yes	Yes	Yes
Note:									*p<0.1; **p<0.05; ***p<0.01

Table 2: Regression results

by imperfect competition on both, labor and product markets, more precisely within efficiency wages, monopsonistic and job-search models. On the contrary negative relation is presented within the standard textbook theory for instance in the work of [17],[2] or [1].

Our results presents a possible explanation for different effects found in literature. The minimum wage growth rates distributions (see Figure 1b) indicate that changes under auto-adjustment mechanism are typically close to zero and hence very smooth ($\overline{wage}_{auto} = -0.003$, $s_{auto} = 0.035$, $\text{skewness}_{auto} = -3.060$). On the other hand there is much greater variability ($\overline{wage}_{state} = 0.009$, $s_{state} = 0.066$) in the *state* sample distribution which is positively skewed ($\text{skewness}_{state} = 0.744$). The positive skew indicates greater tendency to higher positive growth rate in minimum wage growth. These higher changes might create shocks in the labor market which translate in higher unemployment. Unexpected hikes in the level of minimum wage caused by government decisions seem to be significantly disturbing at least in the short run. This happens due to the labor market rigidities and consequent sluggish responses.

4 Conclusion

The effect of the minimum wage is a frequent research topic of the labor economics. However it does not take into account the way of minimum wage setting regime. We distinguish three types of minimum wage setting mechanism: collective bargaining, government unilateral decision with auto adjustment elements and government unilateral decision without any auto adaptation. Our results indicates that effects of minimum wage growth differ with the setting mechanism. The changes under collective bargaining have no significant effect on unemployment which does not hold for the rest. Greater convergence of minimum to median wage leads to decrease in unemployment when the minimum wage is set by government unilateral decision with auto adjustment elements. The effect in case of absence of such elements is precisely opposite.

The effects found are consistent with various theories as well as empirical studies which do not consider the setting mechanism. In fact the finding of different effect under different setting mechanism might explain often opposite impact of minimum wage described by many authors.

Acknowledgements

This research was supported by the Student Project Grant No. MUNI/A/0844/2013.

References

[1] Baker, Michael, Dwayne Benjamin, and Shuchita Stanger: The highs and lows of the minimum wage effect: A time-series cross-section study of the canadian law. *Journal of Labor Economics*, **2** (1999), 318–350.

[2] Bazen, Stephen and Velayoudom Marimoutou: Looking for a needle in a haystack? a re-examination of the time series relationship between teenage employment and minimum wages in the united states*. *Oxford Bulletin of Economics and Statistics*, **64** (supplement) (2002), 699–725.

[3] Benassi, Chiara: *The implementation of minimum wage: Challenges and creative solutions.* ILO, 2011.

[4] Boeri, Tito: Setting the minimum wage. *Labour Economics*, **19** (2012), 281–290.

[5] Brown, Charles: Minimum wage laws: are they overrated? *The Journal of Economic Perspectives*, 1988, p. 133–145.

[6] Calmfors, Lars, John Driffill, Seppo Honkapohja, and Francesco Giavazzi: Bargaining structure, corporatism and macroeconomic performance. *Economic policy*, 1988, p. 14–61.

[7] Card, David: *Myth and measurement: The new economics of the minimum wage.* Princeton University Press, 1995.

[8] Croissant, Yves, and Giovanni Millo: Panel Data Econometrics in `R`: The `plm` Package. *Journal of Statistical Software*, **27** (2008).

[9] Cubitt, Robin P and Shaun P Hargreaves Heap: Minimum wage legislation, investment and human capital. *Scottish Journal of Political Economy*, **46** (1999), 135–157.

[10] Dinkelman, Taryn and Vimal Ranchhod: Evidence on the impact of minimum wage laws in an informal sector: Domestic workers in south africa. *Journal of Development Economics*, **99** (2012), 27–45.

[11] Flinn, Christopher J.: Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. *Econometrica*, **74** (2006), 1013–1062.

[12] Hlavac, Marek: *stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables.* R package version 5.0., (2014).

[13] Katz, Lawrence F. and Alan B. Krueger: The effect of the minimum wage on the fast-food industry. *Industrial and Labor Relations Review*, **46** (1992), 6–21.

[14] Lang, Kevin and Shulamit Kahn: The effect of minimum-wage laws on the distribution of employment: theory and evidence. *Journal of Public Economics*, **69** (1998), 67–82.

[15] Levin-Waldman, Oren M.: The minimum wage and regional wage structure: implications for income distribution. *Journal of economic issues*, (2002), 635–657.

[16] Machin, Stephen and Alan Manning: The effects of minimum wages on wage dispersion and employment: Evidence from the U.K. Wages Councils. *Industrial and Labor Relations Review*, **47** (1994), 319–329.

[17] Neumark, David and William Wascher: Employment effects of minimum and subminimum wages: Panel data on state minimum wage laws. *Industrial and Labor Relations Review*, **46** (1992), 55–81.

[18] Neumark, David and William Wascher: *Minimum wages.* MIT Press, 2008.

[19] R Core Team: *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, 2014.

[20] Ragacs, Christian: *Minimum Wages and Employment: Static and Dynamic Non-market-clearing Equilibrium Models.* Palgrave Macmillan, 2004.

[21] Rebitzer, James B. and Lowell J Taylor: The consequences of minimum wage laws some new theoretical ideas. *Journal of Public Economics*, **56** (1995), 245–255.

[22] Singell, Larry D. and James R. Terborg: Employment Effects Of Two Northwest Minimum Wage Initiatives. *Economic Inquiry*, **45** (2007), 40–55.

[23] Stigler, George J.: The economics of minimum wage legislation. *The American Economic Review*, (1946), 358–365.

The effect of the NRL indicator on the accuracy of financial series forecasts

Monika Miśkiewicz-Nawrocka[1], Katarzyna Zeug-Żebro[2]

Abstract. The real time series consist of a deterministic part and the random factor. The source of random noise may be errors of measurement and errors of rounding made during data analysis. Methods of forecasting time series derived from the theory of nonlinear dynamical systems are sensitive to the presence of random noise. Accordingly, the first stage of the time series prediction should be the data filtration. One of the methods of random noise reduction used for this purpose is the nearest neighbors method.
To assess the level of noise reduction in the filtered time series, an indicator of noise reduction level (NRL) or its modified version can be used. Both of these measures allow to select a series which is characterized by the lowest level of noise after the noise reduction. The question arises whether the most accurate forecast is obtained for the time series of the lowest level of noise.
Since the presence of noise in the data can significantly affect the quality of forecasts, the aim of the article is to evaluate the accuracy of predicting the time series filtered using the nearest neighbor method. The paper will also examine the relation between the indicator NRL and the forecast errors. The test will be conducted on the basis of the financial time series.

Keywords: random noise reduction, nearest neighbor method, financial time series forecasting.

JEL Classification: C01, C53, G17
AMS Classification: 37M10

1 Introduction

Real time series (x_t) are usually disturbed by random noise, therefore (x_t) can be divided into deterministic ($\bar{x}_t$) and stochastic (ξ_t) parts and written in an additive form [10]:

$$x_t = \bar{x}_t + \xi_t , \tag{1}$$

where x_t - observation of the time series at the moment *t*, $\bar{x}_t$ - deterministic part of the time series, ξ_t - stochastic part of the time series which describes the level of random noise in the time series.

The presence of noise in data can significantly affect the quality of forecasting, therefore the first stage of the time series prediction should be the filtration of time series. In literature we can find many procedures used to reduce noise levels, including the method of maximum likelihood [4], the shadowing method [5] and the nearest neighbor method [6]. One of the benefits arising from the use of these methods is the improvement in the forecasting of time series. The effectiveness of the used method of filtration can be tested using the coefficient of noise reduction NRL [10] or its modified version [11]. Both indicators allow for the selection of a series which is characterized by the lowest level of noise after the noise reduction.

The aim of the paper will be to evaluate the accuracy of the prediction of the time series filtered using the nearest neighbor method. The paper will also examine the relation between the NRL indicator (its modified version) and forecast errors. In the study we used the financial time series[3] which were set up with logarithms of daily returns of closing price indices of the world's stock exchanges. The data cover the period from 3 January 2000 to 30 April 2014. The calculations will be carried out using programs written by the authors in the Delphi programming language, Microsoft Excel, Visual Recurrence Analysis 4.2 and TISEAN.

[1] University of Economics in Katowice, Department of Mathematics, monika.miskiewicz@ue.katowice.pl.
[2] University of Economics in Katowice, Department of Mathematics, katarzyna.zeug-zebro@ue.katowice.pl.
[3] The data come from the archive files of website stooq.com.

2 Indicator of the level of random noise (NRL)

One of the procedures used to reduce the level of random noise in the time series is the nearest neighbor method developed by Kantz and Schreiber in 1997[6]. It uses the delay vectors, the so-called *d-stories*, that are the result of the reconstruction of the state space of a dynamic system with the method of delays [7, 13, 15].These vectors take the form of:

$$x_t^d = \left(x_t, x_{t+\tau}, ..., x_{t+(d-1)\tau}\right), \tag{2}$$

where d is the embedding dimension, τ is a natural number, called the delay time, and the variable t takes $n = N - (d-1)\tau$ the value , N is the length of the time series.

The deterministic value of $\bar{x}_l$ for fixed l is determined based on the designated k nearest neighbors of points:

$$_* x_t^d = \left(x_t, x_{t+1}, ..., x_{t+(d-1)}\right), \tag{3}$$

whose one of the central coordinates is the filtered observation x_l (e.g. vector $x^d_{l-\frac{d}{2}}$ for the even value of the embedding dimension d or $x^d_{l-\frac{d+1}{2}}$ for odd values of d), as the arithmetic average of the first coordinates of k nearest neighbors, according to the following formula:

$$\bar{x}_l = \frac{1}{k}\sum_{i=1}^{k} x_{v_i}, \tag{4}$$

where x_{v_i} - are first coordinate of i-th nearest neighbor $x^d_{v_i}$

The effectiveness of the used method of filtration can be tested using the coefficient of noise reduction NRL [10, 16]. It examines the relationship between the strength of the noise added to the system and the geometric structure of its attractor [14]. This indicator is given by the following formula [10]:

$$NRL(d) = \frac{1}{n}\sum_{i=1}^{n} m_i \Big/ \frac{1}{n}\sum_{i=1}^{n} M_i, \tag{5}$$

where m_i and M_i mean the distance from the i-th vector of delays to its nearest and farthest neighbor. The procedure of noise reduction in the time series is carried out for several values of the embedding dimension and the ambient radius of vector x_t^d .

Another measure for the assessment of the noise reduction level is modified indicator NRL_2 proposed by Orzeszko [11]. The formula for this indicator consists of two parts:

$$NRL\_2(d) = \frac{d_{\min} - d^0_{\min}}{d^0_{\min}} + \left|\frac{diam - diam^0}{diam^0}\right|. \tag{6}$$

The first of them $d_{\min} = \frac{1}{n}\sum_{i=1}^{n} m_i$ is a measure of thickening, where m_i means the distance from the i-th vector of delays to its nearest neighbor. The second part of this indicator can be interpreted as the level of a distortion, caused by noise reduction: $diam$, $diam^0$ denote maximal distance between delay vectors, before and after noise reduction respectively.

Using the above criteria, we must choose the one for which coefficient NRL and NRL_2 takes the smallest value.

3 The time series forecasting using the nearest neighbor method NNM [3]

The theoretical basis of the nearest neighbor method is the fact that the states of the deterministic system evolve over time in a similar manner. In the case of time series, if we do not know the function f describing the dynamics of the system and we have only a one-dimensional series of observations $\left(x_1, ..., x_N\right)$, we can use the state space reconstruction [7, 13, 15]. In this way, we obtain delay vectors given by formula (2) that correspond to the original vectors of state space. If $x^d_{t_0}$ is the nearest neighbor of point x^d_N, then also $f_T\left(x^d_N\right) \approx f_T\left(x^d_{t_0}\right)$, which

indicates that $x_{N+T} \approx x_{t_0+T}$. Thus, the value of x_{t_0+T} may be taken as a forecast of observation x_{N+T} in the analyzed time series [9].

In the nearest neighbor method, the forecast for (N+1)-th element $\hat{x}_{N+1}$ is estimated as a weighted average of observations x_{i+1}, where the vectors x_i^d are k nearest neighbors of vector x_N^d in the reconstructed d-dimensional state space [2]:

$$\hat{x}_{N+1} = \sum_{i=1}^{k} w_i x_{i+1} . \tag{7}$$

The weights are chosen so that the closer neighbors have a greater impact on the obtained forecast. Accordingly, the weight of the i-th neighbor is estimated by the formula [2]:

$$w_i = \frac{1}{k-1}\left(1 - \frac{d_i}{d_{TOT}}\right), \tag{8}$$

where $d_i = \left\| x_N^d - x_i^d \right\|$ is the distance between vectors x_N^d and x_i^d, $d_{TOT} = \sum_{i=1}^{k} d_i$, $i = 1,2,\ldots,k$.

In order to estimate the weight of the i-th neighbor can use other formulas presented in [10], but the formula (8) allows to get the most accurate forecasts and takes into account the distance of the nearest neighbors of the forecasted point [8].

4 The purpose and conduct of the study

We studied financial time series which were set up with logarithms of daily returns of closing price indices of the following stock exchanges: AEX – the Amsterdam Exchange index, ATH – Athens Stock Exchange index, BEL 20 –the Euronext Brussels index, CAC40 – the Paris Stock Exchange index (CAC), DAX – the German stock index, FTSE250 – the London Stock Exchange index (FTM), IBEX – the Madrit Stock Exchange index, PX – the Prague Stock Exchange index, SAX – the Bratislava Stock Exchange index and WIG – the Warsaw Stock Exchange index (WIG). The data cover the period from 3 January 2000 to 30 April 2014. The length of the analyzed time series allows to obtain reliable results.

The analysis of the above mentioned time series will run through the following steps:

1. Reduction of the random noise level by the nearest neighbor method and the calculation of the coefficient of the reduction of noise level NRL and its modified version NRL_2.
2. Reconstruction of the state space using the method of delays, i.e. construction of the delay vectors.
3. Forecasting time series using the nearest neighbor method NNM.
4. Quality rating of the designated forecasts by the root of the mean squared error RMSE.

The conducted empirical study allowed to reduce the random noise level in the selected time series using the nearest neighbor method[4]. To perform filtration, we determined the value of the delay time $\tau = 1$ and the values of two parameters: the embedding dimension $d = 2,3,4,5,6,8,10,15$; the ambient radius $\rho = 0.001, 0.01, 0.1$.

In order to assess the reduction in the noise level using the nearest neighbor method, we used the measure NRL(i)[5] for $i = 2,3,\ldots,10$ and NRL_2. The table below (Table 1) contains the lowest value of coefficient NRL and NRL_2 calculated for the analyzed time series and the corresponding embedding dimension and the ambient radius.

[4] Reduction in the random noise conducted using the free program TISEAN developed by H. Kantz and T. Schreiber.

[5] In order to calculate the NRL measure the authors wrote a dedicated programme in the Delphi programming language.

Name of time series	Filtration parameters		NRL	Filtration parameters		NRL_2
	d	ρ		d	ρ	
AEX	2	0.1	0.0005	3	0.1	-1.7932
ATH	3	0.1	0.0011	2	0.1	-1.6392
BEL 20	2	0.1	0.0004	2	0.1	-1.8371
CAC 40	2	0.1	0.0006	2	0.1	-1.7971
DAX	2	0.1	0.0007	2	0.1	-1.8416
FTM	3	0.1	0.0005	2	0.1	-1.9328
IBEX	3	0.1	0.0009	3	0.1	-1.6967
PX	3	0.1	0.0008	2	0.1	-1.5063
SAX	4	0.1	0.0008	3	0.1	-1.5996
WIG20	3	0.1	0.0004	3	0.1	-1.8416

Table 1 Values of measures NRL for the filtered time series

In the next step, we reconstructed the state space with the delays method. Using the method based on the analysis of the autocorrelation function – *ACF* [12], we estimated a time delay τ , while using the false nearest neighbor method – *FNN* [1], we calculated the embedding dimension *d* (Table 2).

Name of time series	Time delays	Embedding dimension	Filtered/Selected by NRL		Filtered/Selected by NRL_2	
			Time delays	Embedding dimension	Time delays	Embedding dimension
AEX	2	6	10	2	10	2
ATH	20	8	2	2	2	6
BEL 20	24	6	4	3	4	3
CAC 40	20	8	1	8	1	8
DAX	13	7	8	7	8	7
FTM	9	7	2	2	1	2
IBEX	19	7	4	2	4	2
PX	11	7	2	2	4	7
SAX	13	12	4	2	4	2
WIG20	3	7	2	2	2	2

Table 2 The parameter values of the reconstruction of the state space of the selected financial time series

Then, to determine a forecast with the NNM method, we used the $K = 2(d+1)$ nearest neighbors of the delay vector (using the Euclidean metric) and the appropriate values of parameters of the reconstruction of the state space (Table 2). In order to compare the results, the study was performed twice for the time series before and after filtration (i.e. for the time series obtained for the parameters *d* and ρ listed in Table 1). Figure 1 shows the prediction error RMSE over the entire interval of verification for the forecast horizon 1, 2, …, 10. The symbol *NameSeries_red*1 designates the examined time series after the reduction of random noise by indicator NRL, and the symbol *NameSeries_red*2 - time series after the reduction of random noise by NRL_2. If time series filtered by NRL and NRL_2 are the same we used the symbol *NameSeries_red*.

Analyzing the obtained results (Figure 1), we can see that the *ex*-post error RMSE obtained for the time series in which the reduction of noise was used are significantly lower than for the forecasts obtained for the original time series. So, we can conclude that the reduction of random noise allowed to improve the accuracy of the obtained forecasts.

Comparing the results obtained for time series chosen/selected by NRL and NRL_2, we can notice that the prediction errors obtained for filtered time series by both measures are comparable. Moreover, for a few time series the measures NRL and NRL_2 indicate exactly the same filtered time series.

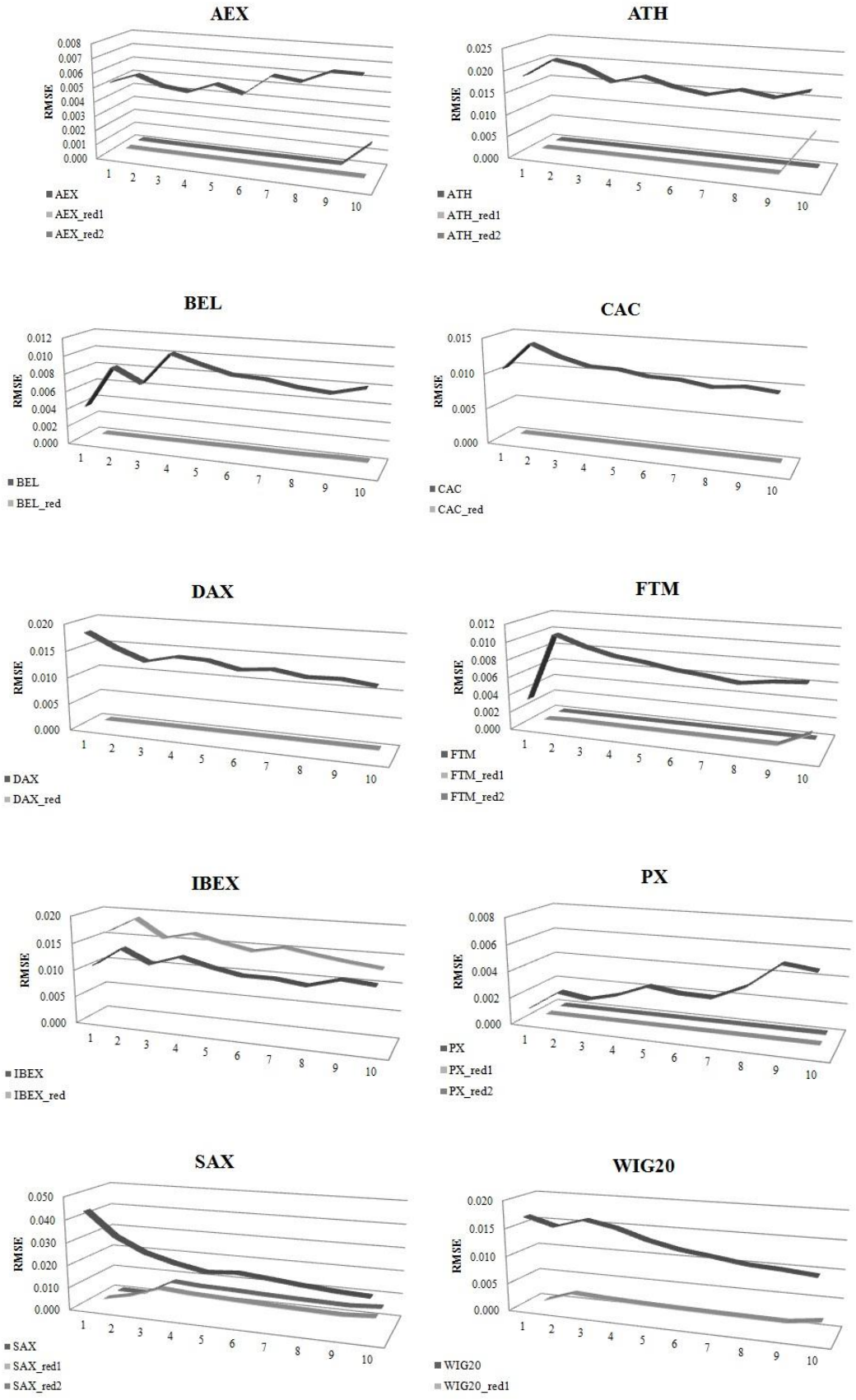


Figure 1 The forecast error RMSE received with the use of the method of nearest neighbors for financial time series for horizon prediction 1, 2, …,10.

5 Conclusions

In the paper we studied the effect of random noise reduction with the nearest neighbor method on the accuracy of forecasts of selected financial time series. Based on the results, we can conclude that for the analyzed time series the ex-post forecast errors obtained for the series that used noise reduction are much lower than the forecasts obtained for the unfiltered series.

In addition, based on the selected financial time series, we compared two selection methods using measures NRL and NRL_2. The conducted study showed that both methods were effective and gave comparable results.

Although the obtained results are rather satisfactory, the reduction of the noise level should not be treated uncritically, because such filtering of data can cause the deformation of the analyzed signal and lead to the misinterpretation of results.

References

[1] Abarbanel, H. D., Brown, R., Kennel, M. B.: *Determining Embedding Dimension for Phase Space Reconstruction Using a Geometrical Construction*. "Physical Review A", 45(6), 1992, 3404-3411.

[2] Chun, S.H., Kim, K.J., Kim, S.H.: *Chaotic analysis of predictability versus knowledge discovery techniques: case study of the Polish stock market*. Expert Systems, Vol. 19, No. 5, 2002, 264 – 272.

[3] Diebold, F. X., Nason, J. A.: *Nonparametric Exchange Rate Prediction?* "Journal of International Economics", 28, 1990, 315 – 332.

[4] Farmer, J. D., Sidorovich, J. J: Exploiting Chaos to Predict the Future and reduce Noise. In: *Evolution, Learning and Cognition*, (ed Y. C. Lee), World Scientific, 1988, 277 – 330.

[5] Farmer, J. D., Sidorovich, J. J.: *Optimal shadowing and noise reduction*, Physica D 47, 1991, 373 – 392.

[6] Kantz, H., Schreiber, T.: *Nonlinear time series analysis*. Cambridge University Press, Cambridge, 1997.

[7] Miśkiewicz-Nawrocka, M.: *Zastosowanie wykładników Lapunowa do analizy ekonomicznych szeregów czasowych*. Uniwersytet Ekonomiczny w Katowicach, Katowice, 2012.

[8] Miśkiewicz-Nawrocka, M.: *The application of random noise reduction by nearest neighbor method to forecasting of economic time series*. Folia Oeconomica Stetinensia (in print).

[9] Nowiński, M.: *Nieliniowa dynamika szeregów czasowych*. Uniwersytet Ekonomiczny we Wrocławiu , Wrocław, 2007.

[10] Orzeszko, W.: *Identyfikacja i prognozowanie chaosu deterministycznego w ekonomicznych szeregach czasowych* . Polskie Towarzystwo Ekonomiczne, Warszawa, 2005.

[11] Orzeszko, W.: *The new method of measuring the effects of noise reduction in chaotic data*. Chaos, Solitons and Fractals, 38, 2008, 1355-1368.

[12] Ramsey, J.B., Sayers, C.L., Rothman, P.: *The Statistical Properties of Dimension Calculations Using Small Data Sets: Some Economic Applications*. International Economic Review, 31, no. 4, 1990.

[13] Takens, F.: Detecting strange attractors in turbulence. In: *Lecture Notes in Mathematics*. (Rand, D. A. and Young, L. S., eds.), Springer, Berlin, 1981, 366 – 381.

[14] Zawadzki, H.: *Chaotyczne systemy dynamiczne*. Uniwersytet Ekonomiczny w Katowicach, Katowice, 1996.

[15] Zeug – Żebro, K.: *Wpływ doboru metod wyznaczania parametrów rekonstrukcji przestrzeni stanów układu dynamicznego na dokładność prognoz*. Zeszyty Naukowe UE Kraków, 898, 2012, 41-54.

[16] Zeug – Żebro, K.: Redukcja szumu losowego metodą najbliższych sąsiadów. (In) *Metody matematyczne, ekonometryczne i komputerowe w finansach i ubezpieczeniach*, (Barczak, A. S., Iskra, D., eds.), Uniwersytet Ekonomiczny w Katowicach, Katowice, 2012, 457-471.

Fuzzy grey AHP

Karel Mls[1], Martin Gavalec[2]

Abstract. Uncertainty as a universal characteristic of the real world problems is discussed in the paper. Analytic Hierarchy Process is one of the most utilized methods for supporting rational decisions. Despite of respected successes of the method, several modifications of the original Saaty's version have been introduced over time.
In the paper, new approach to consideration of information uncertainty - Fuzzy Grey Analytic Hierarchy Process - is introduced. Replacing crisp values in expert comparisons with grey numbers, new ways of improving the quality and accuracy of the model, as well as consensus reaching in group and conflicting decisions are becoming possible.

Keywords: Decision making, AHP, fuzzy logic, grey systems.

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

Incompleteness of information is the general characteristic we have to consider when solving problems and tasks in real world. Moreover, every information is relative and temporal; new information is continually coming to our mind, completing, improving, or changing our understanding of particular problem. Decision making is the main task not only for human experts, for common people, but increasingly it is also conducted by artificial software and hardware systems. Risk and uncertainty are inseparable parts of decision making processes in individual, group or human-machine hybrid teams activities. In these conditions we have to analyze the situation, make many evaluations and select the best available material, service, or supplier.

The scope of the paper is to introduce a novel methodology based on the Analytic Hierarchy Process and Grey System Theory, which enables to describe and manage different levels and forms of information uncertainty and experts' hesitancy in the description and evaluation of the decision model components.

2 Analytic Hierarchy Process and its modifications

Analytic Hierarchy Process (AHP) is a method for formulating and analyzing decisions. Saaty [10] provided a theoretical foundation for the AHP, a decision support tool which can be used to solve complex decision problems taking into account tangible and intangible aspects. The method supports decision makers to make decisions involving their experience, knowledge and intuition. Individual judgments about items, separately on each of several properties are given in the form of paired comparisons w_i/w_j, and relative measures of items are computed. The matrix of paired comparisons A is positive and reciprocal, e.g. $A = (a_{ij}),\ a_{ij} = 1/a_{ji},\ i,j = 1,\dots,n$ and the equation $Aw = nw$, where $w = (w_1, w_2, \dots, w_n)^{\mathrm{T}}$ is the vector of weights, holds. In practical decision making, experts are usually not able to give the precise values of the w_i/w_j, but only their estimations. The matrix A may not be consistent in this case, but is still reciprocal. Computing the vector w then leads to an eigenvalue problem $Aw = \lambda_{max} w$, where λ_{max} is the principal eigenvalue of A, $\lambda_{max} \geq n$. If $\lambda_{max} = n$, then the matrix A is consistent; for $\lambda_{max} > n$ several methods were proposed to express the level of matrix A inconsistency. Saaty has shown [9], that the consistency index $\mathrm{CI} = (\lambda_{max} - n)/(n-1)$ divided by the same index obtained as an average over a large number of random reciprocal matrices of the same order gives the comparable results and we can say, that if the ratio $\mathrm{CR} = \mathrm{CI}/\mathrm{CI}_{rand} \lesssim 0.1$, the inconsistency of the comparison matrix A is acceptable.

Besides the eigenvector method, several other mathematical methods has been proposed to find the relevant weight vector from known matrix of pair comparisons in more simple way so far [1]. In our paper, geometric row means g method will be considered.

[1] University of Hradec Králové, Faculty of Informatics and Management, Department of Information Technologies, Rokitanského 62,Hradec Králové, Czech Republic, e-mail: karel.mls@uhk.cz.
[2] University of Hradec Králové, Faculty of Informatics and Management, Department of Information Technologies, Rokitanského 62,Hradec Králové, Czech Republic, e-mail: martin.gavalec@uhk.cz.

$$g_i = (\prod_{m=1}^{n} a_{im})^{1/n}, \quad (1)$$

2.1 Fuzzy AHP

Pairwise comparisons, the main characteristic of the AHP, require decision-maker's preferences in the form of a precise evaluation of alternatives or criteria proportions by the 1-9 fundamental scale. In practice, human thinking and reasoning is usually fuzzy, considering the possibility of partial membership of the element in a set more acceptable than the crisp option. Criteria *Size of house* and *Modern facilities* from the example discussed in the section 4 are typical representatives of a fuzzy set.

Formal fuzzification of the AHP (FAHP) has been introduced by Laarhoven and Pedrycz [3] soon after the original Saaty's method [10]. In fact, fuzzy approach in some form (of linguistic variables) is used to evaluate the relative importance of the criteria and to rate alternatives with respect to criteria [12]. The usual approach is based on the transformation of real elements in the comparison matrix into the fuzzy numbers or into fuzzy linguistic variables. Then, after specific fuzzy operations, alternatives' rating is obtained. Specific field of FAHP analysis is dedicated to the consistency of the fuzzy model [8].

2.2 Interval AHP

In general, precise values or comparisons of elements in multi criteria model cannot be obtained and only estimations are available. This uncertainty in the information can be addressed by a decision maker through the use of an interval judgment rather than a single numerical value [7]. The width of the interval is then a measure of the uncertainty of the individual judgment and can be either minimized by obtaining more relevant information about the system, or it can be utilized for minimizing conflicts and different points of view in group decision making situations. By interval AHP the priority weights of alternatives and importance weights of decision makers are obtained as interval from the given pairwise comparison matrices. The resulting interval priority weights reflect all possibilities in the given information and can help the group members to understand one another and reach the consensus [2].

3 Grey numbers and systems

Grey System Theory (GST) has been designed for solving high uncertainty problems with small and incomplete discrete data sets [5]. GST includes fuzziness, moreover, fuzzy mathematics needs some previous information (usually based on cognitive experiences); while GST handles objective data, it does not require any previous information other than the data to be disposed. Main characteristics and differences between GST, probability and fuzzy mathematics are shown in Table 1.

	Grey System Theory	**Probability statistics**	**Fuzzy mathematics**
Objects of study	Poor information uncertainty	Stochastic uncertainty	Cognitive uncertainty
Basic sets	Grey hazy sets	Cantor sets	Fuzzy sets
Methods	Information coverage	Probability distribution	Function of affiliation
Procedure	Grey series generation	Frequency distribution	Marginal sampling
Requirement	Any distribution	Typical distribution	Experience
Emphasis	Intension	Intension	Extension
Objective	Laws of reality	Laws of statistics	Cognitive expression
Characteristics	Small samples	Large samples	Experience

Table 1 Comparison between Grey Systems Theory, probability statistics and fuzzy mathematics

Fundamental principles of grey systems according to [6]:

- *principle of information difference* (difference implies the existence of information. Each piece of information must carry some kind of difference);
- *principle of non-uniqueness* (the solution to any problem with incomplete and nondeterministic information is not unique);
- *principle of minimal information* (making the most and best use of the available 'minimal amount of information');

- *principle of recognition base* (information is the foundation on which people recognize and understand nature);
- *principle of new information priority* (the function of new pieces of information is greater than that of old pieces of information);
- *principle of absoluteness of greyness* (incompleteness of information is absolute).

A grey number is the basic concept of the GST. The exact value of the grey number is unknown, but only a range within its value lies, together with additional information about the range. We can assume several classes of grey numbers:

Grey numbers with only a lower limit

$$\otimes G \in [\underline{G}, \infty), \tag{2}$$

where $\underline{G}$ is the lower limit of the grey number $\otimes G$ and $[\underline{G}, \infty)$ the value field of the grey number $\otimes G$.

Grey numbers with only an upper limit

$$\otimes G \in (-\infty, \overline{G}], \tag{3}$$

where $\overline{G}$ is the upper limit of the grey number $\otimes G$.

Interval grey numbers

A grey number with both lower limit and upper limit is called interval grey number, $\otimes G \in [\underline{G}, \overline{G}]$. If both limits are fixed numbers, we call the $\otimes G$ first order interval grey number. Another kind of grey numbers are characterized by a base value a and vibration variable δ_a , so $\otimes G(a) \in [a - \delta_a, a + \delta_a]$.

Continuous and discrete grey numbers

The grey numbers with a finite or a countable number of values in a given interval are called discrete grey numbers. Those grey numbers whose values cover an interval are called continuous grey numbers.

Black and white numbers

When the grey number has neither upper nor lower limit, or the upper and the lower limits are grey numbers, $\otimes G \in (-\infty, \infty)$ or $\otimes G \in (\otimes G_1, \otimes G_2)$, then $\otimes G$ is called black number.

When $\otimes G \in [\underline{G}, \overline{G}]$ and $\underline{G} = \overline{G}$, then $\otimes G$ is called white number.

Operations on grey numbers

$$\otimes G_1 + \otimes G_2 \in [\underline{G}_1 + \underline{G}_2, \overline{G}_1 + \overline{G}_2], \tag{4}$$

$$\otimes G_1 - \otimes G_2 \in [\underline{G}_1 - \overline{G}_2, \overline{G}_1 - \underline{G}_2], \tag{5}$$

$$\otimes G_1 \cdot \otimes G_2 \in \left[min\left(\underline{G}_1 \cdot \underline{G}_2, \underline{G}_1 \cdot \overline{G}_2, \overline{G}_1 \cdot \underline{G}_1, \overline{G}_1 \cdot \overline{G}_2\right), max\left(\underline{G}_1 \cdot \underline{G}_2, \underline{G}_1 \cdot \overline{G}_2, \overline{G}_1 \cdot \underline{G}_1, \overline{G}_1 \cdot \overline{G}_2\right)\right], \tag{6}$$

$$\frac{\otimes G_1}{\otimes G_2} \in \left[min\left(\frac{\underline{G}_1}{\overline{G}_2}, \frac{\underline{G}_1}{\underline{G}_2}, \frac{\overline{G}_1}{\overline{G}_2}, \frac{\overline{G}_1}{\underline{G}_2}\right), max\left(\frac{\underline{G}_1}{\overline{G}_2}, \frac{\underline{G}_1}{\underline{G}_2}, \frac{\overline{G}_1}{\overline{G}_2}, \frac{\overline{G}_1}{\underline{G}_2}\right)\right], \tag{7}$$

$$\lambda \cdot \otimes G \in [\lambda \cdot \underline{G}, \lambda \cdot \overline{G}], \tag{8}$$

Whitenization of grey numbers and degree of greyness

$$\widetilde{\otimes} G = \alpha \cdot \underline{G} + (1 - \alpha) \cdot \overline{G}, \tag{9}$$

where $0 \leq \alpha \leq 1$. For general interval grey number $\otimes G \in [\underline{G}, \overline{G}]$ and $\alpha = 0.5$, the whitenization value $\widetilde{\otimes}$ is called equal mean whitenization.

GST has been studied not only from a theoretical point of view, but there are also many applications. Grey systems benefits in models and tools for better decision making attracted several authors [11, 4].

4 Grey and fuzzy grey approach to AHP

Although the fundamental scale for pairwise comparisons [10] has been carefully designed to enable wide range of (nonlinear) evaluations with numerical and verbal intensities and meanings (from 1 – equal importance to 9 – extreme importance), it only enables to express the *cognitive uncertainty* of the decision maker. To cover also *information uncertainty*, interval, or more general grey systems extension of the standard AHP method is needed.

Grey comparison matrix $\otimes A$ is composed of the grey numbers representing expert's imperfect knowledge about pairs of elements A_i, A_j of the hierarchy.

$$\otimes A = \begin{bmatrix} 1 & \otimes a_{12} & \dots & \otimes a_{1n} \\ \otimes a_{21} & 1 & \dots & \otimes a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \otimes a_{n1} & \otimes a_{n2} & \dots & 1 \end{bmatrix}, \tag{10}$$

where $\otimes a_{ij} = 1/\otimes a_{ji}$ and $\underline{a}_{ij} = 1/9, \overline{a}_{ij} = 9$.

Then, using (1) and (6) we obtain the local grey priority vector in the form:

$$\otimes w = (\otimes w_1, \otimes w_2, \dots, \otimes w_n)^T, \ \otimes w_i \in \left[\left(\prod_{m=1}^{n} \underline{a}_{im} \right)^{1/n}, \left(\prod_{m=1}^{n} \overline{a}_{im} \right)^{1/n} \right] . \tag{11}$$

Before synthesizing partial results to the global ranking of alternatives, normalization and whitenization of the local grey priority vector is done.

$$\otimes w_{norm} = (\otimes w_{1norm}, \otimes w_{2norm}, \dots, \otimes w_{nnorm})^T, \ \otimes w_{inorm} \in \left[\frac{\underline{w}_i}{\sum_{k=1}^{n} \underline{w}_k}, \frac{\overline{w}_i}{\sum_{k=1}^{n} \overline{w}_k} \right] , \tag{12}$$

$$\widetilde{w} = (\widetilde{w}_{1norm}, \widetilde{w}_{2norm}, \dots, \widetilde{w}_{nnorm})^T, \ \widetilde{w}_{inorm} = \alpha \cdot \underline{w}_{inorm} + (1 - \alpha) \cdot \overline{w}_{inorm} , \tag{13}$$

5 Numerical example

To illustrate the proposed method, we will compare the AHP and FGAHP approach on a simple example, *Choosing the best house to buy*, adopted from [9]. We focus only on one element of the model, *Size of house*, as it represents possible information uncertainty in appropriate way. The full hierarchy of the decision problem is shown in Fig. 1, where the component of our interest is highlighted.

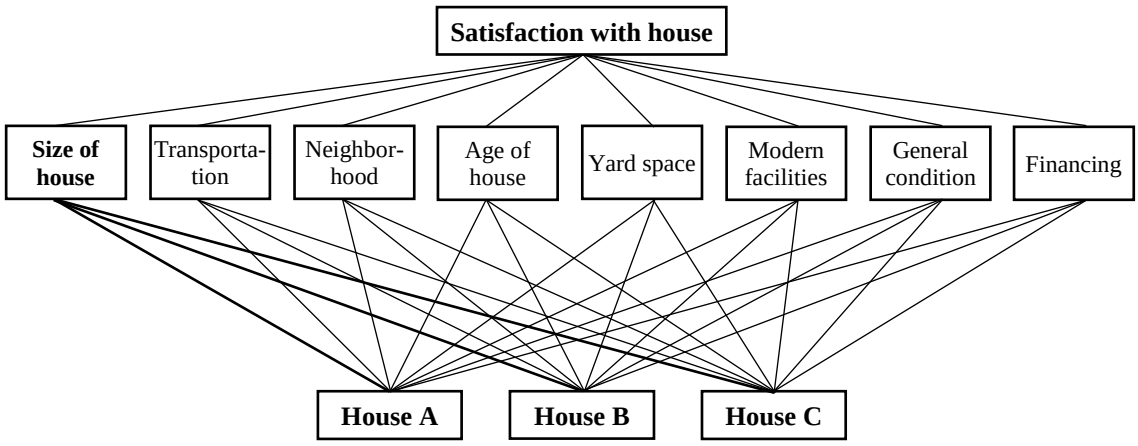


Figure 1 Hierarchical model of the Choosing the best house problem

According to Saaty, House A is the largest of the three alternatives, House B is a little smaller than House A, and House C is very small. The matrix of original comparisons of the houses with respect to the criterion *Size of house* and its local priority is given in Table 2. It is evident, that expert numerical evaluation of the linguistic

characteristics is not in good accordance with the fundamental scale intensities definitions („House B is a little smaller than House A“ – House B/House A = 1/6 = 1/“strong plus“). Such imperfections may lead on one hand to variance of calculated priorities of elements and to unpredictable inconsistency of the model, on the other hand they encourage the use of interval or GST methods.

Size of house	A	B	C	Local priority
House A	1	6	8	0.754
House B	1/6	1	4	0.181
House C	1/8	1/4	1	0.065

Table 2 Comparison matrix and local priority for AHP

Local priority vector from [9] has been compared with values of the same vector obtained by the geometric means method and we found no difference. So the same method has been applied to the grey modification of the model.

Size of house	A	B	C	Grey local priority	Whitenized local priority
House A	1	[5,7]	[6,8]	[0.717, 0.770]	0.743
House B	[1/7,1/5]	1	[3,4]	[0.167, 0.195]	0.180
House C	[1/8,1/6]	[1/4,1/3]	1	[0.063, 0.088]	0.076

Table 3 Comparison matrix and local priority for FGAHP; $\alpha = 0.5$

Now the expert comparisons are holding incomplete information about the component of the model. The uncertainty is expressed by interval grey numbers with lower and upper limits and the grey local priority vector is computed according to (11), (12) and (13). The whitenized local priority vector has slightly different values of its components, compared to original AHP result, which is the consequence of non-symmetric lower and upper limits of grey numbers for the comparisons House A/House C and House B/House C, and the equal mean whitenization method used. However, these differences caused no change in preference ranking of alternatives with respect to the given criterion.

6 Conclusions

New method of AHP with considering incomplete information in evaluation of multicriteria decision model was introduced. For the comparison with some most influential existing approaches, short characteristic of classical AHP, fuzzy AHP and interval AHP was revised. Then, formal approach for crisp comparison matrix extension on grey numbers was proposed. In the numerical example, comparison of original-crisp AHP and proposed fuzzy-grey AHP was illustrated.

Next research will focus on the problems in consistency measuring and evaluating in FGAHP, and on forms of evaluation of group decision making and consensus reaching using the FGAHP method.

Acknowledgements

This research was supported by the grant No. 14-02424S of the Czech Science Foundation (GA ČR).

Many thanks to the anonymous reviewers for their constructive suggestions and comments on this paper.

References

[1] Dijkstra, T.K.: On the extraction of weights from pairwise comparison matrices. *Central European Journal of Operations Research*, **21**(2013), 103-123.

[2] Etani, T.: Interval AHP for group decision makers. In: Proceedings .IFSA/EUSFLAT 2009 Congress, Calouste Gulbenkian Foundation, Lisbon, Portugal, 2009, 155–160.

[3] Laarhoven van, P.J.M., Pedrycz, W.: A fuzzy extension of Saaty’s priority theory. *Fuzzy Sets and Systems* **11**, 4 (1983), 229–241.

[4] Li, G.-D., Yamaguchi, D., Nagai, M.: A grey-based decision-making approach to the supplier selection problem. *Mathematical and Computer Modelling* **46** (2007), 573-581.

[5] Liu, S., Lin, Y.: *Grey Information. Theory and Practical Applications.* Springer/Verlag London, 2006.

[6] Liu, S., Lin, Y.: *Grey Systems. Theory and Applications*. Springer-Verlag, Berlin Heidelberg, 2010.
[7] Moreno-Jimenez, J.M., Vargas, L.G.: A probabilistic study of preference structures in the analytic hierarchy process with interval judgements. *Mathematical and Computer Modelling* **17**, 4/5 (1993), 73-81.
[8] Ramík, J., Korviny, P.: Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean. *Fuzzy Sets and Systems* **161** (2010), 1604-1613.
[9] Saaty, T.L.:How to make a decision: The Analytic Hierarchy Process. *European Journal of Operational Research* **48** (1990), 9-26.
[10] Saaty, T.L.: Continuous, recursive and feedback hierarchic systems. Working paper, University of Pittsburgh, 1981.
[11] Salmeron, J.L., Papageorgiou, E.I.: A Fuzzy Grey Cognitive Maps-based Decision Support System for Radiotherapy Treatment Planning. *Knowledge-Based Systems* **30**, 1 (2012), 151-160.
[12] Torfi, F., Farahani, R.Z., Rezapour, S.: Fuzzy AHP to determine the relative weights of evaluation criteria and Fuzzy TOPSIS to rank the alternatives. *Applied Soft Computing* **10** (2010), 520-528.

Robustness of Monge matrices in fuzzy algebra

M. Molnárová[1]

Abstract. Robust Monge matrices over max-min algebra (fuzzy matrices) are studied. The max-min algebra (fuzzy algebra) is an extremal algebra with operations maximum and minimum. Equivalent conditions for robustness of Monge fuzzy matrices are presented. Polynomial algorithm for checking the necessary and sufficient conditions for Monge fuzzy matrices to be robust is described.

Keywords: (max, min) algebra, robustness, Monge matrix

JEL classification: C02
AMS classification: 08A72, 90B35, 90C47

1 Introduction

Discrete dynamic systems (DDS), graph theory, knowledge engineering or description of technical devices are some of the areas in which the extremal so–called $(\max, \min)$–algebra (with operations maximum and minimum) is useful for modelling applications. The matrices representing a DDS in $(\max, \min)$–algebra are the so–called fuzzy matrices. The Monge matrices and their applications in many different areas were studied in [1], [2]. Properties of fuzzy matrices were described in [3]. The problem of robust matrices in fuzzy algebra was presented in [8]. Sufficient and necessary conditions for checking robustness of interval fuzzy matrices with corresponding polynomial algorithms were introduced in [6]. Robustness of Monge fuzzy matrices in binary case was presented in [5].

The aim of this paper is to describe the robustness of Monge fuzzy matrices and to find polynomial algorithm for verifying the robustness.

2 Background of the problem

The fuzzy algebra $\mathcal{B}$ is a triple $(B, \oplus, \otimes)$, where $(B, \leq)$ is a bounded linearly ordered set with binary operations *maximum* and *minimum*, denoted by $\oplus$, $\otimes$. The least element in B will be denoted by O, the greatest one by I. By $\mathbb{N}$ we denote the set of all natural numbers. The greatest common divisor of a set $S \subseteq \mathbb{N}$ is denoted by $\gcd S$, the least common multiple of the set S is denoted by $\operatorname{lcm} S$. For a given natural $n \in \mathbb{N}$, we use the notation N for the set of all smaller or equal positive natural numbers, i.e., $N = \{1, 2, \ldots, n\}$.

For any $m, n \in \mathbb{N}$, $B(m, n)$ denotes the set of all matrices of type $m \times n$ and $B(n)$ the set of all n-dimensional column vectors over $\mathcal{B}$. The matrix operations over $\mathcal{B}$ are defined formally in the same manner (with respect to $\oplus$, $\otimes$) as matrix operations over any field. The rth power of a matrix $A \in B(n, n)$ is denoted by A^r, with elements a_{ij}^r. For $A, C \in B(n, n)$ we write $A \leq C$ if $a_{ij} \leq c_{ij}$ holds for all $i, j \in N$.

A *digraph* is a pair $G = (V, E)$, where V, the so–called vertex set, is a finite set, and E, the so–called edge set, is a subset of $V \times V$. A digraph $G' = (V', E')$ is a subdigraph of the digraph G (for brevity $G' \subseteq G$), if $V' \subseteq V$ and $E' \subseteq E$. A path in the digraph $G = (V, E)$ is a sequence of vertices $p = (i_1, \ldots, i_{k+1})$ such that $(i_j, i_{j+1}) \in E$ for $j = 1, \ldots, k$. The number k is the length of the path p and is denoted by $\ell(p)$. If $i_1 = i_{k+1}$, then p is called a cycle. For a given matrix $A \in B(n, n)$ the symbol $G(A) = (N, E)$ stands for the complete, edge–weighted digraph associated with A, i.e., the vertex set of $G(A)$ is N, and the capacity of any edge $(i, j) \in E$ is a_{ij}. In addition, for given $h \in B$, the *threshold digraph* $G(A, h)$ is the digraph $G = (N, E')$ with the vertex set N and the edge set $E' = \{(i, j);\ i, j \in N,\ a_{ij} \geq h\}$.

The following lemma describes the relation between matrices and corresponding threshold digraphs.

[1]Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, e-mail: Monika.Molnarova@tuke.sk

Lemma 1. [6] Let A, $C \in B(n,n)$. Let h, h_1, $h_2 \in B$.

(i) If $A \leq C$ then $G(A,h) \subseteq G(C,h)$,

(ii) if $h_1 < h_2$ then $G(A,h_2) \subseteq G(A,h_1)$.

By a *strongly connected component* of a digraph $G(A,h) = (N,E)$ we mean a subdigraph $\mathcal{K} = (N_\mathcal{K}, E_\mathcal{K})$ generated by a non-empty subset $N_\mathcal{K} \subseteq N$ such that any two distinct vertices $i,j \in N_\mathcal{K}$ are contained in a common cycle, $E_\mathcal{K} = E \cap (N_\mathcal{K} \times N_\mathcal{K})$ and $N_\mathcal{K}$ is the maximal subset with this property. A strongly connected component $\mathcal{K}$ of a digraph is called non-trivial, if there is a cycle of positive length in $\mathcal{K}$. For any non-trivial strongly connected component $\mathcal{K}$ is the *period* of $\mathcal{K}$ defined as $\operatorname{per}\mathcal{K} = \gcd\{\,\ell(c);\ c \text{ is a cycle in } \mathcal{K},\ \ell(c) > 0\,\}$. If $\mathcal{K}$ is trivial, then $\operatorname{per}\mathcal{K} = 1$. By $\mathrm{SCC}^\star(G)$ we denote the set of all non-trivial strongly connected components of G.

Let $A \in B(n,n)$ and $x \in B(n)$. The sequence $O(A,x) = \{x^{(0)}, x^{(1)}, x^{(2)}, \dots, x^{(n)}, \dots\}$ is the orbit of $x = x^{(0)}$ generated by A, where $x^{(r)} = A^r \otimes x^{(0)}$ for each $r \in \mathbb{N}$.
For a given matrix $A \in B(n,n)$, the number $\lambda \in B$ and the n–tuple $x \in B(n)$ are the so–called *eigenvalue* of A and *eigenvector* of A, respectively, if they are the solution of the *eigenproblem* for matrix A, i.e. they satisfy the equation $A \otimes x = \lambda \otimes x$. The corresponding *eigenspace* $V(A,\lambda)$ is defined as the set of all eigenvectors of A with associated eigenvalue λ, i.e., $V(A,\lambda) = \{x \in B(n);\ A \otimes x = \lambda \otimes x\}$.
Let $\lambda \in B$. A matrix $A \in B(n,n)$ is *ultimately λ–periodic* if there are natural numbers p and R such that the following holds: $A^{k+p} = \lambda \otimes A^k$ for all $k \geq R$. The smallest natural number p with above property is called the period of A, denoted by $\operatorname{per}(A,\lambda)$. In case $\lambda = I$ we denote $\operatorname{per}(A,I)$ by abbreviation $\operatorname{per} A$.

According to [3] we define

$$\begin{aligned} \mathrm{SCC}^\star(A) &= \cup\{\,\mathrm{SCC}^\star(G(A,h));\ h \in \{a_{ij}; i,j \in N\}\,\}, \\ \mathrm{SCC}^{\min}(A) &= \{\,\mathcal{K} \in \mathrm{SCC}^\star(A);\ \mathcal{K} \text{ is minimal in } \mathrm{SCC}^\star(A), \text{ ordered by inclusion}\,\}. \end{aligned}$$

Theorem 1. [3] *Let $A \in B(n,n)$. Then*

(i) $\operatorname{per} A = \operatorname{lcm}\{\operatorname{per}\mathcal{K};\ \mathcal{K} \in \mathrm{SCC}^\star(A)\}$,

(ii) $\operatorname{per} A = \operatorname{lcm}\{\operatorname{per}\mathcal{K};\ \mathcal{K} \in \mathrm{SCC}^{\min}(A)\}$.

Theorem 2. [3] *There is an algorithm by which, for a given $A \in B(n,n)$*

(i) $\operatorname{per} A$ *can be computed in $O(n^3)$ time,*

(ii) if $\mathrm{SCC}^{\min}(A)$ *is given, then* $\operatorname{per} A$ *can be computed in $O(n^2)$ time.*

Definition 1. Let $A = (a_{ij}) \in B(n,n)$, $\lambda \in B$. Let $T(A,\lambda) = \{x \in B(n);\ O(A,x) \cap V(A,\lambda) \neq \emptyset\}$. A is called *λ–robust* if $T(A,\lambda) = B(n)$. A λ–robust matrix with $\lambda = I$ is called *a robust matrix.*

In our considerations we will use the following result (adapted for $\lambda = I$) proved in [8] to study robustness of a matrix.

Lemma 2. [8] Let $A = (a_{ij}) \in B(n,n)$. Then A is robust if and only if $\operatorname{per} A = 1$.

3 Robustness of Monge matrices

Definition 2. We say, that a matrix $A = (a_{ij}) \in B(m,n)$ is a convex Monge matrix (concave Monge matrix) if and only if

$$\begin{aligned} a_{ij} \otimes a_{kl} &\leq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l \\ \big(\, a_{ij} \otimes a_{kl} &\geq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l\,\big). \end{aligned}$$

In this paper, we assume that the considered matrix is convex.

Theorem 3. [5] *Let $A \in B(n,n)$ be a (0-1) matrix with no zero rows and no zero columns. A is a Monge matrix if and only if $a_{ij} = 0$ implies*

$$\textit{either} \qquad a_{kl} = 0 \qquad \textit{for all} \qquad k \leq i,\ l \leq j \tag{1}$$

$$\textit{or} \qquad a_{kl} = 0 \qquad \textit{for all} \qquad k \geq i,\ l \geq j. \tag{2}$$

It is not enough to check two consecutive rows and columns to verify the Monge property of a matrix in max-min algebra as illustrated in the next example. Hence the computational complexity for checking the Monge property of a fuzzy matrix is $O(n^3)$ ([1], [2]).

Example 1. For the matrix

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 1 & 4 \end{pmatrix}$$

the consecutive columns satisfy the Monge property. Unfortunately $a_{11} \otimes a_{23} > a_{13} \otimes a_{21}$.

Lemma 3. [4] Let $c = (i_1, i_2, \ldots, i_k, i_1)$ be a cycle of length $k \geq 3$. Then there are arcs (i_j, i_{j+1}) and (i_l, i_{l+1}) in c such that

$$i_j < i_l \quad \text{and} \quad i_{j+1} < i_{l+1}.$$

Let us denote by $h^{(1)}, h^{(2)}, \ldots, h^{(r)}$ the elements of the set $H = \{a_{ij};\, i, j \in N\}$ ordered into a strictly decreasing sequence, i.e.,

$$h^{(1)} > h^{(2)} > \cdots > h^{(r)}. \tag{3}$$

The number r is equal to the number of different values in the matrix A.

Lemma 4. Let $A \in B(n, n)$. Then the sequence of threshold digraphs corresponding to the sequence (3) is ordered by inclusion

$$G(A, h^{(1)}) \subseteq G(A, h^{(2)}) \subseteq \ldots \subseteq G(A, h^{(r)}).$$

Proof. The assertion follows by Lemma 1. □

Lemma 5. Let $A \in B(n, n)$ be a Monge matrix. Let $h \in H$. Let $\mathcal{K} \in \mathrm{SCC}^\star(G(A, h))$. Let c be a cycle of odd length $\ell(c) \geq 3$ in $\mathcal{K}$. Then there is a node in c with a loop.

Proof. Let $h \in H$. Let $\mathcal{K} \in \mathrm{SCC}^\star(G(A, h))$. Let $c = (i_1, i_2, \ldots, i_k, i_1)$ be a cycle of odd length $\ell(c) \geq 3$ in $\mathcal{K}$. By Lemma 3 there are arcs (i_j, i_{j+1}) and (i_l, i_{l+1}) in c such that $i_j < i_l$ and $i_{j+1} < i_{l+1}$. By the Monge property

$$h \leq a_{i_j i_{j+1}} \otimes a_{i_l i_{l+1}} \leq a_{i_l i_{j+1}} \otimes a_{i_j i_{l+1}}.$$

Thus $h \leq a_{i_l i_{j+1}}$ and $h \leq a_{i_j i_{l+1}}$. Consequently the component $\mathcal{K}$ contains arcs (i_j, i_{l+1}) and (i_l, i_{j+1}) as well. Hence the cycle c splits into cycles $c_1 = (i_1, i_2, \ldots, i_j, i_{l+1}, \ldots, i_1)$ and $c_2 = (i_{j+1}, \ldots, i_l, i_{j+1})$ with $\ell(c) = \ell(c_1) + \ell(c_2)$. In addition exactly one of these cycles is of odd length. If the length of this cycle is one the assertion follows. If the length is more than one, we repeat the procedure and split the cycle with odd length into two cycles. This procedure will be repeated until we get a cycle of length one. □

Lemma 6. Let $A \in B(n, n)$ be a Monge matrix. Let $h \in H$. Let for $i, k \in N$ be the loops (i, i) and (k, k) in the digraph $G(A, h)$. Then the nodes i and k are in the same non-trivial strongly connected component $\mathcal{K}$ of $G(A, h)$.

Proof. Let $h \in H$. Let $i, k \in N$ with loops (i, i) and (k, k) in $G(A, h)$. Then $a_{ii} \geq h$ and $a_{kk} \geq h$. If $i = k$ the assertion trivially follows. Let us assume that $i \neq k$. By the Monge property

$$h \leq a_{ii} \otimes a_{kk} \leq a_{ik} \otimes a_{ki}.$$

Hence $h \leq a_{ik}$ and $h \leq a_{ki}$. Consequently there is a common cycle for nodes i and k in $G(A, h)$. Thus i and k are in the same non-trivial strongly connected component $\mathcal{K} \in G(A, h)$. □

Theorem 4. *Let $A \in B(n, n)$ be a Monge matrix. Then A is robust if and only if for each $h \in H$ the digraph $G(A, h)$ contains at most one non-trivial strongly connected component and this has a loop.*

Proof. Let $A \in B(n, n)$ be a Monge matrix. Let us assume that A is robust. Let there be $h \in H$ such that $G(A, h)$ contains two distinct non-trivial strongly connected components $\mathcal{K}_1$ and $\mathcal{K}_2$ each with a loop or $G(A, h)$ contains one non-trivial strongly connected component without any loop.

Case 1. Let $(i, i) \in \mathcal{K}_1$ and $(k, k) \in \mathcal{K}_2$ be the loops in considered non-trivial strongly connected components $\mathcal{K}_1$ and $\mathcal{K}_2$ of $G(A, h)$. Due to Lemma 6 the nodes i and k lie in the same non-trivial strongly connected component. What is a contradiction.

Case 2. Let $G(A,h)$ contains exactly one non-trivial strongly connected component $\mathcal{K}$ with no loop. Hence there is no cycle of odd length with $l(c) \geq 3$ in $\mathcal{K}$ by Lemma 5. Moreover, since $\mathcal{K}$ is a non-trivial component every cycle is of even length in $\mathcal{K}$. Thus $\operatorname{per}\mathcal{K} > 1$. Consequently by Theorem 1 $\operatorname{per} A > 1$. What is a contradiction.

For the converse implication let us assume that $G(A,h)$ contains at most one non-trivial strongly connected component with a loop for each $h \in H$. Due to Theorem 1 and Lemma 2 the matrix A is robust. □

Example 2. Let us check the robustness of the given Monge matrix $A \in B(6,6)$ for $B = [0,10]$

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 3 & 2 \\ 0 & 0 & 4 & 4 & 2 & 0 \\ 0 & 3 & 4 & 3 & 0 & 0 \\ 0 & 3 & 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Due to Theorem 4 we shall verify that $G(A,h)$ contains at most one non-trivial strongly connected component and this with a loop for each $h \in H = \{0,1,2,3,4\}$. Since both, component $\mathcal{K}_1$ generated by the node set $N_{\mathcal{K}_1} = \{2,5\}$ and $\mathcal{K}_2$ generated by the node set $N_{\mathcal{K}_2} = \{3,4\}$ are non-trivial strongly connected components of $G(A,3)$ (see Figure 1) the considered matrix is not robust.

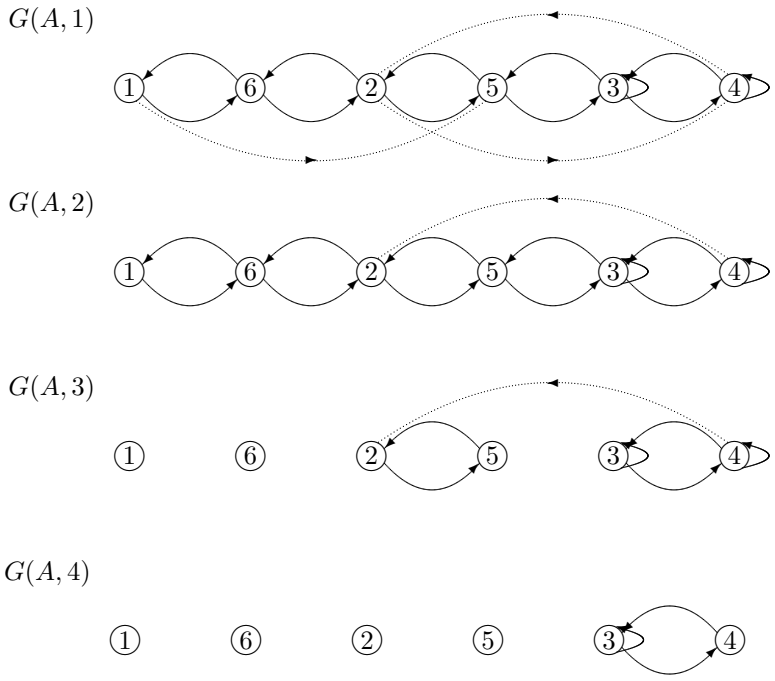


Figure 1: Threshold digraphs in non–robust case

The matrix C in the next example is a slight modification of the above matrix A but the answer is positive.

Example 3. Let us check the robustness of the given Monge matrix $C \in B(6,6)$ which arises from the above matrix A for $B = [0,10]$. Modified elements are highlighted by bold characters.

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 3 & 2 \\ 0 & 0 & 4 & 4 & \mathbf{3} & 0 \\ 0 & 3 & 4 & 3 & 0 & 0 \\ 0 & 3 & \mathbf{4} & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

In contrast to the previous example the digraph $G(C,h)$ contains at most one non-trivial strongly connected component and this with a loop for each $h \in H = \{0,1,2,3,4\}$ (see Figure 2). Due to Theorem 4 the considered matrix is robust.

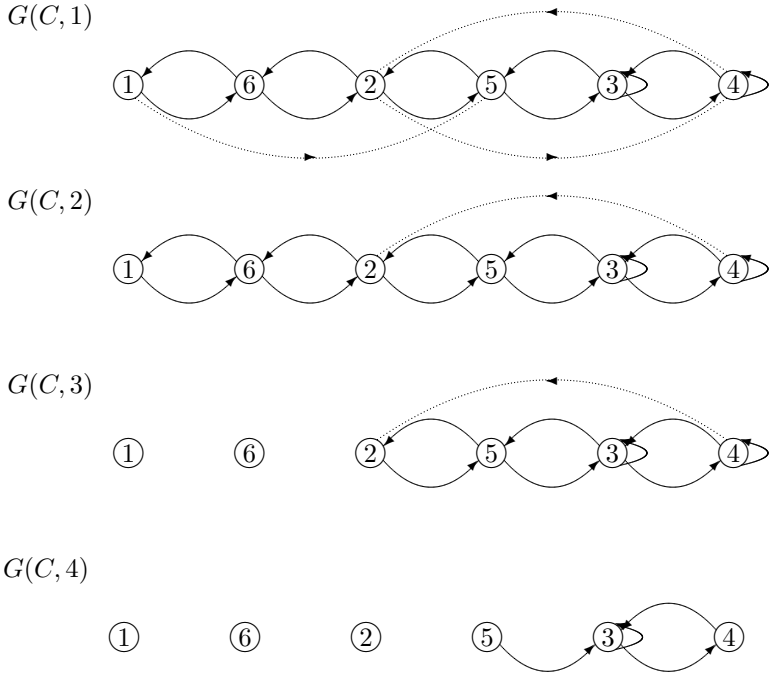


Figure 2: Threshold digraphs in robust case

In view of Theorem 1 we can formulate even a stronger result in regard to relation between the non-trivial strongly connected components corresponding to the sequence of thresholds (3).

Theorem 5. *Let $A \in B(n,n)$ be a Monge matrix. Then A is robust if and only if one of the statements holds*

(i) $\mathrm{SCC}^{\min}(A) = \emptyset$,

(ii) $\mathrm{SCC}^{\min}(A) = \{\mathcal{K}\}$ *and* $\mathcal{K}$ *contains a loop.*

Proof. Let $A \in B(n,n)$ be a Monge matrix. Let us assume that A is robust. Due to Theorem 4 there are two possibilities.

Case 1. $G(A,h)$ contains no non-trivial strongly connected component for any $h \in H$. Then $\mathrm{SCC}^{\star}(A) = \emptyset$. Consequently $\mathrm{SCC}^{\min}(A) = \emptyset$.

Case 2. Let there be $h^{(\star)} \in H$ such that $G(A, h^{(\star)})$ contains a non-trivial strongly connected component $\mathcal{K}_{h^{(\star)}}$. By Theorem 4 this is the only non-trivial strongly connected component of $G(A, h^{(\star)})$, moreover, with a loop. In addition there exists $h^{(max)} \geq h^{(\star)}$ such that $h^{(max)}$ is the maximum value in H with the property that the corresponding digraph $G(A, h^{(max)})$ contains a non-trivial strongly connected component $\mathcal{K}_{h^{(max)}}$ and this with a loop. Consequently $G(A, h^{(max)}) \subseteq G(A, h^{(\star)})$ by Lemma 1. Furthermore, using Theorem 4 for all $h \in H$ satisfying $h^{(r)} \leq h \leq h^{(max)}$ the digraph $G(A,h)$ contains exactly one non-trivial strongly connected component $\mathcal{K}_h$ and this with a loop. Due to Lemma 4 we obtain the following ordered sequence of non-trivial strongly connected components

$$\mathcal{K}_{h^{(max)}} \subseteq \ldots \subseteq \mathcal{K}_{h^{(r-1)}} \subseteq \mathcal{K}_{h^{(r)}}$$

corresponding to the sequence $h^{(max)} > \cdots > h^{(r-1)} > h^{(r)}$. Hence $\mathcal{K}_{h^{(max)}}$ is the only minimal strongly connected component of $\mathrm{SCC}^{\star}(A)$ and contains a loop. Thus $\mathrm{SCC}^{\min}(A) = \{\mathcal{K}_{h^{(max)}}\}$.

The converse implication is a consequence of Theorem 1 and Lemma 2. □

It is not necessary to compute the period of a Monge matrix to decide about the robustness. Hence we do not need the Balcer-Veinott algorithm. Unfortunately, to find the set of minimal components requires the computation of the metric matrix by Floyd-Warshall algorithm with computational complexity $O(n^3)$ ([3]). We recall that also the verification of the Monge property takes $O(n^3)$ time.

Theorem 6. *There is an algorithm by which, for a given Monge matrix $A \in B(n,n)$*

(i) the robustness can be verified in $O(n^3)$ time,

(ii) if $\mathrm{SCC}^{\min}(A)$ is given, then the robustness can be verified in $O(1)$ time.

Proof. (ii) If $\mathrm{SCC}^{\min}(A)$ is given, we check robustness of A in $O(1)$ time due to the following algorithm:

Algorithm **Robustness**

Input. $\mathrm{SCC}^{\min}(A)$.
Output. 'yes' in variable *rob*, if A is robust;

'no' in variable *rob*, if A is not robust.

begin

(i) **If** $|\mathrm{SCC}^{\min}(A)| = 0$, **then** *rob* :='yes'; **else go to** step (ii);

(ii) **If** $|\mathrm{SCC}^{\min}(A)| > 1$, **then** *rob* :='no'; **else go to** step (iii);

(iii) **If** $\mathrm{SCC}^{\min}(A) = \{\mathcal{K}\}$ and $\mathcal{K}$ contains a loop, **then** *rob* :='yes'; **else** *rob* :='no';

end

(i) First we find the set $\mathrm{SCC}^{\min}(A)$ by the algorithm described in [3] in $O(n^3)$ time. Second we decide about robustness of A in $O(1)$ time by algorithm **Robustness**. □

Example 4. Let us check the robustness of above matrices A and C in view of Theorem 5.

According to the definition of the set $\mathrm{SCC}^{\min}(A)$ until $h = 3$ the digraph $G(A,h)$ contains no minimal strongly connected component. There are two non-trivial strongly connected components in $G(A,3)$ (see Figure 1). Namely, $\mathcal{K}_1$ generated by the node set $N_{\mathcal{K}_1} = \{2,5\}$ and $\mathcal{K}_2$ generated by the node set $N_{\mathcal{K}_2} = \{3,4\}$. Since the digraph $G(A,4)$ contains a non-trivial strongly connected component $\mathcal{K}_2'$ generated by the node set $N_{\mathcal{K}_2'} = \{3,4\}$ and $\mathcal{K}_2' \subset \mathcal{K}_2$ the component $\mathcal{K}_2$ is not minimal. Consequently $\mathrm{SCC}^{\min}(A) = \{\mathcal{K}_1, \mathcal{K}_2'\}$ and by Theorem 5 the matrix A is not robust.

For the matrix C the set $\mathrm{SCC}^{\min}(C)$ consists of exactly one non-trivial strongly connected component, namely, $\mathcal{K}_2'$ which is with a loop (see Figure 2). Thus the matrix C is robust.

Acknowledgements

This work was supported by the Slovak Cultural and Education Agency No. 032 TUKE- 4/2013.

References

[1] Burkard, R. E.: Monge properties, discrete convexity and applications, *EJOR*, Volume **176** (2007), 1–14.

[2] Burkard, R. E., Klinz, B. and Rudolf, R.: Perspectives of Monge properties in optimization, *DAM*, Volume **70** (1996), 95–161.

[3] Gavalec, M.: *Periodicity in extremal algebras.* GAUDEAMUS, Hradec Králové, 2004.

[4] Gavalec, M. and Plavka, J.: An $O(n^2)$ algorithm for maximum cycle mean of Monge matrices in max algebra, *DAM*, Volume **127** (2003), 651–656.

[5] Molnárová, M.: Robustness of (0–1) Monge fuzzy matrices, In: *Proceedings of* 31^{st} *Int. Conference Mathematical Methods in Economics 2013*, Jihlava, 2013, 636–641.

[6] Molnárová, M., Myšková, H. and Plavka, J.: The robustness of interval fuzzy matrices, *DAM*, Volume **438** (2013), 3350–3364.

[7] Myšková, H. and Štefanský, L.: Robustness of fuzzy interval circulant–Hankel matrices, *LAA*, Volume **444** (2014), 165–182.

[8] Plavka, J. and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *DAM*, Volume **159** Issue 5 (2011), 381–388.

Interval Max-drast Systems of Linear Equations

Helena Myšková[1]

Abstract.

Max-drast algebra uses instead of conventional operations for addition and multiplication of vectors and matrices the operations maximum and one of the triangular norms, the drastic norm, respectively. Transition matrices in max-drast algebra and their power sequences occur in describing complex fuzzy systems in which extreme demands are put on the reliability of the system.

Many practical problems lead to solving systems of linear equations, which are given by transition matrix A and right-hand side vector b of suitable size. However, in practice we deal often with inexact input data. This leads to the requirement to replace the scalar matrices and vectors with so-called interval matrices and vectors. We can define several types of solvability of interval systems. In this paper, we shall deal with the weak, control and tolerance solvability of interval max-drast systems of linear equations. We give necessary and sufficient conditions for each of them.

Keywords: max-drast algebra, interval system, weak solvability, tolerance solvability, control solvability.

JEL classification: C02
AMS classification: 15A18; 15A80; 65G30

1 Introduction

Discrete time systems are often described with transition matrices with fuzzy values. In dependence on the character of the system, various triangular fuzzy t-norms are used. If the elements of a transition matrix correspond to the flow capacities, then the behavior of such system is described by powers of matrices in max-min algebra. Periodic behavior of the power sequence in max-min algebra was studied by many authors, see e. g. [5], [6], in max-plus algebra in [1] and in max-drast algebra in [7].

The drastic triangular norm is the basic example of a non-divisible t-norm on any partially ordered set, see [2]. Max-drast matrix concerns the transitions reliability in the system. Interpretation of a complex fuzzy systems in max-drast fuzzy algebra reflects the situation when extreme demands are put on the reliability of the system. The problems of the solvability of interval systems in max-plus and max-min algebra has been studied in [3, 8, 9].

In this paper we derive necessary and sufficient conditions for the solvability concepts of interval systems of linear equations in max-drast algebra.

2 Preliminaries

By *max-drast algebra* we understand a triplet $(\mathcal{I}, \oplus, \otimes_d)$, where $\mathcal{I}$ is the closed unit interval $[0, 1]$, $\oplus$ is the maximum and $\otimes_d$ is the binary drast operation (drastic triangular norm) defined as follows

$$x \otimes_d y = \begin{cases} \min(x, y) & \text{if } \max(x, y) = 1, \\ 0 & \text{if } \max(x, y) < 1. \end{cases} \tag{1}$$

Proposition 1. *Let $a \in \mathcal{I}$ be arbitrary. Then $a \otimes_d 0 = 0$ and $a \otimes_d 1 = a$.*

[1]Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovak Republic, helena.myskova@tuke.sk

Proposition 1 says that number 1 is a neutral element with respect to $\otimes_d$ and 0 is an absorbing element with respect to $\otimes_d$. The drast operation is commutative, associative and distributive due to the operation of maximum. Similarly as max-plus and max-min algebra, max-drast algebra is an extremal algebra as defined in [12].

By $\mathbb{N}$ we denote the set of all natural numbers. For given natural numbers $n, m \in \mathbb{N}$, we use the notations $N = \{1, 2, \ldots, n\}$ and $M = \{1, 2, \ldots, m\}$. For any $n, m \in \mathbb{N}$, $\mathcal{I}(m, n)$ denotes the set of all $m \times n$ matrices and $\mathcal{I}(n)$ denotes the set of all n-dimensional column vectors over $\mathcal{I}$. The matrix operations over $\mathcal{I}$ are defined formally in the same manner (with respect to $\oplus$, $\otimes_d$) as matrix operations over any field.

For $b = (b_i) \in \mathcal{I}(n)$, $d = (d_i) \in \mathcal{I}(n)$ we write $b \leq d$ if $b_i \leq d_i$ holds for all $i \in N$. Similarly, we write $A \leq C$ for $A = (a_{ij}) \in \mathcal{I}(m, n)$, $C = (c_{ij}) \in \mathcal{I}(m, n)$ if $a_{ij} \leq c_{ij}$ holds for all $i \in M$, $j \in N$.

3 Systems of linear equations in max-drast algebra

Each system of linear equations in max-drast algebra can be written in the matrix form

$$A \otimes_d x = b, \tag{2}$$

where $A \in \mathcal{I}(m, n)$ and $b \in \mathcal{I}(m)$. The solvability of (2) in max-plus algebra and fuzzy algebra was studied by many authors, see e. g. [4], [12].

Denote by ξ the hypothetical number infinitely close to number 1, but less than 1. This follows that $\min\{x, \xi\} = x$ for each $x \in \mathcal{I} - \{1\}$ and $\min\{1, \xi\} = \xi$. To give a necessary and sufficient condition for the solvability of (2) we define the operation $\otimes'_d$ as follows:

$$a \otimes'_d b = \begin{cases} b & \text{if } a = 1, \\ \xi & \text{if } a > b,\ a \neq 1, \\ 1 & \text{if } a \leq b. \end{cases} \tag{3}$$

In max-drast algebra, we define a *principal solution* $x^*(A, b)$ as follows:

$$x^*_j(A, b) = \min_{i \in M} \{a_{ij} \otimes'_d b_i\}, \quad j \in N. \tag{4}$$

An importance of the principal solution for the solvability of the system of max-drast linear equations is expressed by the following assertions, which are similar to the assertions in max-plus and max-min algebra, see [4], [12].

Lemma 1. Let $A \in B(m, n)$ and $b \in B(m)$ be given. Then

(i) if $A \otimes_d x = b$ for some $x \in B(n)$, then $x \leq x^*(A, b)$,

(ii) $A \otimes_d x^*(A, b) \leq b$.

Proof.

(i) Suppose that a vector $x \in \mathcal{I}(n)$ is such that $A \otimes_d x = b$. We shall prove that $x_j \leq x^*_j(A, b)$ for each $j \in N$. If $x^*_j(A, b) \in \{\xi, 1\}$ then $x_j \leq x^*_j(A, b)$ trivially holds. Suppose that $x^*_j(A, b) = b_i < 1$. Then $a_{ij} = 1$ and the equality $[A \otimes_d x]_i = b_i$ implies $x_j = a_{ij} \otimes_d x_j \leq b_i = x^*_j(A, b)$.

(ii) We have to prove that $[A \otimes_d x^*(A, b)]_i \leq b_i$ for each $i \in M$, i. e. $a_{ij} \otimes_d x^*_j(A, b) \leq b_i$ for each $i \in M$, $j \in N$. For the sake of contradiction suppose that there exist $i \in M$, $j \in N$ such that $a_{ij} \otimes_d x^*_j(A, b) > b_i$. There are two possibilities.

If $a_{ij} = 1$ and $x^*_j(A, b) > b_i$ then we have a contradiction with the definition of $x^*(A, b)$.

If $a_{ij} > b_i$ and $x^*_j(A, b) = 1$ then the equality $x^*_j(A, b) = 1$ implies $a_{ij} \leq b_i$ for each $i \in M$, a contradiction. □

Remark 1. We will say that (2) is solvable if there exists $x \in \mathcal{I}(n)$ (a solution of (2)) such that $A \otimes_d x = b$ and $x_j \neq \xi$ for each $j \in N$.

Theorem 2. *Let $A \in B(m,n)$ and $b \in B(m)$ be given. Then the system of linear equations $A \otimes_d x = b$ is solvable if and only if $A \otimes_d x^*(A,b) = b$.*

Proof. Suppose that the system of linear equations $A \otimes_d x = b$ is solvable, i. e., there exists $y \in \mathcal{I}(n)$ such that $A \otimes_d y = b$. According to Lemma 1 (i) we have $y \leq x^*(A,b)$ and consequently $A \otimes_d x^*(A,b) \geq A \otimes_d y = b$. From Lemma 1 (ii) we obtain $b \geq A \otimes_d x^*(A,b) \geq b$, i.e., $A \otimes_d x^*(A,b) = b$.

For the converse implication suppose that $A \otimes_d x^*(A,b) = b$. If $x_j^*(A,b) \neq \xi$ for each $j \in N$, then $x^*(A,b)$ is a solution of (2). If $x_j^*(A,b) = \xi$ for some $j \in N$ then $a_{ij} \neq 1$ and consequently $a_{ij} \otimes x_j^*(A,b) = 0$, for each $i \in M$. Thus we get a solution of (2) by replacing all ξ-s with arbitrary numbers from interval $[0,1)$. □

To give another necessary and sufficient condition for the solvability of (2) we denote

$$N^* = \{j \in N;\, x_j^*(A,b) \neq \xi\}, \quad M_j = \{i \in M;\, a_{ij} \otimes_d x_j^*(A,b) = b_i\} \text{ for each } j \in N^*.$$

Theorem 3. *Let $A \in B(m,n)$ and $b \in B(m)$ be given. Then the system of linear equations $A \otimes_d x = b$ is solvable if and only if $\bigcup\limits_{j \in N^*} M_j = M$.*

Proof. Let $\bigcup\limits_{j \in N^*} M_j = M$ and $i \in M$ be arbitrary. Then there exists $j \in N^*$ such that $a_{ij} \otimes_d x_j^*(A,b) = b_i$. According to Lemma 1 (ii) we get $[A \otimes x^*(A,b)]_i = b_i$. Thus $A \otimes x^*(A,b) = b$ and in the case that $x_j^*(A,b) = \xi$ for some $j \in N$ we get a solution of (2) by the same considerations as in the proof of Theorem 2.

For the converse implication suppose that there exists $y \in \mathcal{I}(n)$ such that $A \otimes_d y = b$, i.e. $[A \otimes_d y]_i = b_i$ for each $i \in M$. Let $i \in M$ be arbitrary, but fixed. Then there exists $j \in N$ such that $a_{ij} \otimes_d y_j = b_i$. The inequalities $y \leq x^*(A,b)$, $A \otimes_d x^*(A,b) \leq b$ imply $a_{ij} \otimes_d x_j^*(A,b) = b_i$. From (3) and (4) it follows that $x_j^*(A,b) \neq \xi$, so $j \in N^*$. Therefore for each $i \in M$ there exists $j \in N^*$ such that $a_{ij} \otimes_d x_j^*(A,b) = b_i$, i. e., $i \in M_j$. Thus $\bigcup\limits_{j \in N^*} M_j = M$. □

Lemma 2. A system of inequalities

$$A \otimes_d x \leq b, \tag{5}$$

$$B \otimes_d x \geq c \tag{6}$$

has a solution if and only if $B \otimes_d x^*(A,b) \geq c$.

Proof. Suppose that there exists $y \in \mathcal{I}(n)$ such that $A \otimes_d y \leq b$, $B \otimes_d y \geq c$. Then $y \leq x^*(A,b)$ which implies $B \otimes_d x^*(A,b) \geq B \otimes_d y \geq c$.

For the converse implication it suffices to replace all ξ-s in $x^*(A,b)(A,b)$ with the value equal to $\max\{c_i : [B \otimes_d x^*(A,b)]_i = \xi\}$. The obtained vector is a solution of (5) according to Lemma 1. The validity of (6) follows from the assumption. □

.

4 Solvability of interval systems

In practice, the elements in a a matrix and right-hand side vector may depend on outside conditions, so the values a_{ij} are from intervals of possible values. In this section we shall deal with matrices and vectors with interval elements. Similarly to [8], [9], [11] we define an interval matrix $\boldsymbol{A}$ and an interval vector $\boldsymbol{b}$.

Definition 1. Let $\underline{A}, \overline{A} \in \mathcal{I}(m,n)$, $\underline{A} \leq \overline{A}$ and $\underline{b}, \overline{b} \in \mathcal{I}(m)$, $\underline{b} \leq \overline{b}$. An interval matrix $\boldsymbol{A}$ and an interval vector $\boldsymbol{b}$ are defined as follows

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \mathcal{I}(m,n);\, \underline{A} \leq A \leq \overline{A} \right\},$$

$$\boldsymbol{b} = [\underline{b}, \overline{b}] = \left\{ b \in \mathcal{I}(m);\, \underline{b} \leq b \leq \overline{b} \right\}.$$

An *interval system of linear equations* is a system of the form

$$\boldsymbol{A} \otimes_d x = \boldsymbol{b}, \tag{7}$$

where $\boldsymbol{A} = [\underline{A}, \overline{A}]$ is an interval matrix with $\underline{A}, \overline{A} \in B(m,n)$, $\underline{A} \leq \overline{A}$ and $\boldsymbol{b} = [\underline{b}, \overline{b}]$ is an interval vector with $\underline{b}, \overline{b} \in B(m)$, $\underline{b} \leq \overline{b}$. In fact, each interval system (7) represents a set of systems of the form (2) such that $A \in \boldsymbol{A}$ and $b \in \boldsymbol{b}$. Each such system is called a *subsystem* of interval system (7).

We shall think over the solvability of interval system on the ground of solvability of its subsystems. In the following, we define several types of solvability of max-drast interval systems of linear equations.

4.1 Weak solvability

The weak solvability of an interval system requires the solvability of at least one subsystem. In max-plus algebra, the weak solvability was studied in [3], but the same assertions hold in max-min algebra.

First, we define the notion of a possible solution.

Definition 2. A vector x is a possible solution of interval system (7) if there exist $A \in \boldsymbol{A}$ and $b \in \boldsymbol{b}$ such that $A \otimes_d x = b$.

Theorem 4. *A vector x is a possible solution of interval system* (7) *if and only if it satisfies the system of inequalities*

$$\underline{A} \otimes_d x \leq \overline{b}, \tag{8}$$

$$\overline{A} \otimes_d x \geq \underline{b}. \tag{9}$$

Proof. Let $x \in \mathcal{I}(n)$ be given. Then the product $[A \otimes_d x]_i$ as a function of variables $a_{i1}, a_{i2}, \ldots, a_{in}$. This is an isotonic continuous function and so the image of the n-dimensional interval $[\underline{a}_{i1}, \overline{a}_{i1}] \times [\underline{a}_{i2}, \overline{a}_{i2}] \times \cdots \times [\underline{a}_{in}, \overline{a}_{in}]$ is the interval $[[\underline{A} \otimes_d x]_i, [\overline{A} \otimes_d x]_i]$. The existence of $A \in \boldsymbol{A}$ and $b \in \boldsymbol{b}$ such that $A \otimes_d x = b$ means that $[[\underline{A} \otimes_d x]_i, [\overline{A} \otimes_d x]_i] \cap [\underline{b}_i, \overline{b}_i] \neq \emptyset$ for each $i \in M$, which is equivalent to the system of inequalities (8), (9). □

Definition 3. Interval system (7) is weakly solvable if there exist $x \in \mathcal{I}(n)$, $A \in \mathcal{I}(m,n)$ and $b \in \mathcal{I}(m)$ such that $A \otimes_d x = b$.

Theorem 5. *Interval system* (7) *is weakly solvable if and only if*

$$\overline{A} \otimes_d x^*(\underline{A}, \overline{b}) \geq \underline{b}. \tag{10}$$

Proof. The weak solvability of (7) means that there exists a possible solution $x \in \mathcal{I}(n)$ which is according to Theorem 4 equivalent to the solvability of the system of inequalities (8), (9). By Lemma 2 we get (10). □

4.2 Tolerance solvability

In [3], the author dealt with the tolerance solvability of max-plus interval systems. The published assertions hold also in max-min algebra.

Definition 4. A vector x is a tolerance solution of interval system (7) if for each $A \in \boldsymbol{A}$ there exists $b \in \boldsymbol{b}$ such that $A \otimes_d x = b$.

Theorem 6. *A vector x is a tolerance solution of interval system* (7) *if and only if it satisfies the system of inequalities*

$$\overline{A} \otimes_d x \leq \overline{b}, \tag{11}$$

$$\underline{A} \otimes_d x \geq \underline{b}. \tag{12}$$

Proof. A vector x is a tolerance solution of (7), if for each $A \in \boldsymbol{A}$ the product $A \otimes_d x$ lies in $\boldsymbol{b}$. This leads to the requirement for the validity of the system of inequalities $\underline{b} \leq A \otimes_d x \leq \overline{b}$ for each $A \in \boldsymbol{A}$. The left inequality is satisfied for each $A \in \boldsymbol{A}$ if and only if $\underline{A} \otimes_d x \geq \underline{b}$, i. e. inequality (12) holds, and the right one is equivalent to (11). □

Definition 5. Interval system (7) is tolerance solvable if there exist $x \in \mathcal{I}(n)$ such that for each $A \in \boldsymbol{A}$ there exists $b \in \boldsymbol{b}$ such that $A \otimes_d x = b$.

Theorem 7. *Interval system* (7) *is tolerance solvable if and only if*

$$\underline{A} \otimes_d x^*(\overline{A}, \overline{b}) \geq \underline{b}. \tag{13}$$

Proof. The tolerance solvability of (7) is equivalent to the existence of a vector x such that x is a tolerance solution. The existence of a tolerance solution is by Theorem 6 equivalent to the solvability of system of inequalities (11), (12), which is according to Lemma 2 equivalent to (13). □

4.3 Control solvability

The control solvability in max-plus and max-min algebra was studied in [9]. In this solvability concept, the existence of a control solution, defined below, is required.

Definition 6. A vector x is a control solution of interval system (7) if for each $b \in \boldsymbol{b}$ there exists $A \in \boldsymbol{A}$ such that $A \otimes_d x = b$.

Theorem 8. *A vector x is a control solution of interval system* (7) *if and only if it satisfies the system of inequalities*

$$\underline{A} \otimes_d x \leq \underline{b}, \tag{14}$$

$$\overline{A} \otimes_d x \geq \overline{b}. \tag{15}$$

Proof. A vector x is a control solution if and only if $\boldsymbol{b} \subseteq \{A \otimes_d x;\ A \in \boldsymbol{A}\}$, i.e., $[\underline{b}_i, \overline{b}_i] \subseteq [[\underline{A} \otimes_d x]_i, [\overline{A} \otimes_d x]_i]$ for each $i \in M$, which is equivalent to the system of inequalities (14), (15). □

Definition 7. Interval system (7) is control solvable if there exists $x \in \mathcal{I}(n)$ such that for each $b \in \boldsymbol{b}$ there exists $A \in \boldsymbol{A}$ such that $A \otimes_d x = b$.

Theorem 9. *Interval system* (7) *is control solvable if and only if*

$$\overline{A} \otimes x^*(\underline{A}, \underline{b}) \geq \overline{b}. \tag{16}$$

Proof. Interval system (7) is control solvable if and only if it has a control solution. In view of Theorem 8 the existence of a control solution is equivalent to the solvability of the system of inequalities (14), (15). Using Lemma 2 we get inequality (16). □

4.4 The relationship between solvability concepts

In the previous, we defined three solvability concepts for max-drast interval system of linear equations. The relationship between them is expressed by the following proposition.

Proposition 10. *Let an interval max-drast system of linear equations in the form (7) be given. The following assertions hold:*

(i) *If* (7) *is tolerance solvable, then it is weakly solvable.*

(ii) *If* (7) *is control solvable, then it is weakly solvable.*

The proof follows immediately from the definition of the mentioned solvability concepts.

Example 1. We check whether the given interval system $\boldsymbol{A} \otimes_d x = \boldsymbol{b}$ is weakly, tolerance and control solvable, if

$$\boldsymbol{A} = \begin{pmatrix} [1,1] & [0.5,0.7] & [0.3,1] \\ [0.4,0.4] & [1,1] & [0.3,0.8] \\ [0.5,0.7] & [0.7,0.9] & [1,1] \end{pmatrix} \text{ and } \boldsymbol{b} = \begin{pmatrix} [0.4,0.5] \\ [0.6,1] \\ [0.5,0.7] \end{pmatrix}.$$

First, we check the weak solvability. We compute $x_1^*(\underline{A},\overline{b}) = \min\{0.5,\xi,\xi\} = 0.5$, $x_2^*(\underline{A},\overline{b}) = \min\{1,1,1\} = 1$, $x_3^*(\underline{A},\overline{b}) = \min\{1,1,0.7\} = 0.7$.

We have $x^*(\underline{A},\overline{b}) = (0.5,1,0.7)^\top$ and $\overline{A} \otimes_d x^*(\underline{A},\overline{b}) = (0.7,1,0.9)^\top \geq (0.4,0.6,0.5) = \underline{b}$. Since inequality (10) is satisfied, according to Theorem 5 the given interval system is weakly solvable. Moreover, the possible solution is $x = x^*(\underline{A},\overline{b}) = (0.5,1,0.7)^\top$. (it may not be the only).

For checking the tolerance solvability we compute the vector $x^*(\overline{A},\overline{b}) = (0.5,\xi,0.5)^\top$ and the product $\underline{A} \otimes_d x^*(\overline{A},\overline{b}) = (0.5,\xi,0.5)^\top$. Since $\underline{A} \otimes_d x^*(\overline{A},\overline{b}) \geq \underline{b}$, the given interval system is tolerance solvable. To find a tolerance solution it suffices, according to Lemma 2, to replace ξ in $x_2^*(\overline{A},\overline{b})$ with $\max\{\underline{b}_i : [\underline{A} \otimes_d x^*(\overline{A},\overline{b})]_i = \xi\} = \underline{b}_2 = 0.6$. So the vector $x = (0.5,0.6,0.5)^\top$ is a tolerance solution of the given interval system.

Finally we verify the control solvability. We have $x^*(\underline{A},\underline{b}) = (0.4,0.6,0.5)^\top$. Since $\overline{A} \otimes_d x^*(\underline{A},\underline{b}) = (0.5,0.6,0.5)^\top \not\geq \overline{b}$, the given interval system is not control solvable.

References

[1] Butkovič, P. and Cuninghame-Green, R. A.: On matrix powers in max-algebra. *Lin. Alg. Appl.* **421** (2007) 370–381.

[2] Cassanovas, J., and Mayor, G.: Discrete t-norms and operations on extented multisets, *Fuzzy sets and systems* **159** (2008), 1165–1177.

[3] K. Cechlárová, *Solutions of interval systems in max-plus algebra*, Proc.SOR 2001, Eds.V. Rupnik, L.Zadnik-stirn, S.Drobne, Preddvor, Slovenia, 321-326.

[4] R. A. Cuninghame-Green, *Minimax Algebra*, Lecture notes in Economics and Mathematical systems **1966**, Springer, Berlin, 1979.

[5] Gavalec, M.: Computing matrix period in max-min algebra, *Discrete Appl. Math.* **75** (1997), 63–70.

[6] Gavalec, M.: Computing orbit period in fuzzy algebra, *Discrete Appl. Math.* **100** (2000), 49–65.

[7] Gavalec, M., and Němcová, Z.: Matrix period in max-drast fuzzy algebra. In: *Proceedings of 30th International Conference Mathematical Methods in Industry*, Karviná, 2012, 213–218.

[8] Myšková, H.: Interval systems of max-separable linear equations, *Linear Alg. Appl.* 403 (2005), 263 –272.

[9] Myšková, H.: Control solvability of interval systems of max-separable linear equations, *Lin. Alg. Appl.* **416** (2006), 215 –223.

[10] Plavka, J., and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *Discrete Appl. Math.* **159** no. 5 (2011), 381–388.

[11] Plavka, J.: On the $O(n^3)$ algorithm for checking the strong robustness of interval fuzzy matrices, *Discrete Applied Math.* **160** (2012), 640-647.

[12] K. Zimmermann, *Extremální algebra*, Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV, Praha, 1976.

Influence of Technical Parameters of Underpasses and Economical Aspects on Wildlife Migrations

Kateřina Myšková [1], Jaroslav Žák [2]

Abstract. Ecological impacts of line structures are increasingly discussed. Underpasses are understood as an important element in ensuring the permeability of motorways. Width, height and length are the main technical parameters that affect the migration of animals. At first glance it is clear that large values of the width and height of the underpass ensure its good functionality, but cause high financial costs. Small values of the underpass length are appropriate in terms of both ecology and economics, but are not realistic. The aim of the article is based on the data using relevant statistical methods to ascertain whether the technical parameters affect the migration of animals and how. A logit model can be an appropriate statistical tool in this case. Authors have got a large amount of data for different types of underpasses obtained using camera systems, spy games and traditional forms of monitoring. Medium and large mammals were mainly objects of interest. Furthermore, the economic impacts of different types of underpasses have been examining.

Keywords: Wildlife migration, underpasses, economic analysis, logit model.

JEL classification: C89, Q51
AMS classification: 62J12

1 Introduction

Wildlife underpasses are designed to increase the probability of permeability of line traffic structures (freeways, highways) for feral species. Insufficient permeability of the highway for feral animals can lead to fragmentation of populations and dividing territory into small parts which do not allow long time survival of the species [6], [13] and [20]. Environmentalists sometime tend to demand construction of expensive objects even in places where it is not necessary. Investors logically try to reduce the cost of the unnecessary precautions.

The most economical way how to ensure necessary permeability of highways for feral animals is to use current bridges over watercourses or communications crossing the highway. Should this be insufficient there is a need to design special migration objects (underpasses, overpasses) which increase the cost. Correct proposal of migration object has to include very precise estimation of migration potential. Unnecessary increase of costs or decrease of migration potential might happen if the estimation is imprecise. Previously mentioned principles are further described in [3], [4] or [9].

Methodology for calculation of migration potential was taken from [15] and [20]. The migration potential is calculated as the product of the ecological and the technical migration potential which is based on technical parameters of underpasses, specifically height, width and length. Meaning of the formulations width, height and length of underpasses is explained in Figure 1. We should note that the span or length of the bridge for structural engineers means width of the underpass for environmentalists.

There are quite reliable data available concerning the behaviour of the feral spieces [1], [11] and their spread on certain territory [5], [7], [14]. But there are only few papers about the evaluation of the functionality of underpasses based on statistical procedures and video records [8], [16], [17], [19]. Costbenefit analyses of mitigation measures are also available [2], [10], [20], but do not cover financial analysis of structures versus migration potential. While monitoring of the functionality of wildlife overpasses has

[1]Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, myskova@mendelu.cz

[2]The Institute of Technology and Businesses in České Budějovice, Department of Civil Engineering, Okružní 517/10, 370 01 České Budějovice, Czech Republic, jzax@seznam.cz

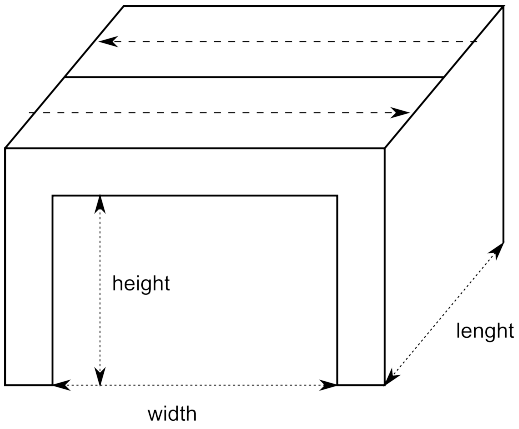


Figure 1 Illustration of underpass' size

become quite frequent in many countries [4], [11], [12], empirical studies that deal with population-level effects and investigate whether general guidelines can be developer are very rare [18].

Purpose of this article is to analyze by statistical methods whether technical parameters really affect permeability of migration object and in which way. We assume that migration permeability will grow with increasing width and height. The permeability will drop with increasing length of the underpass.

2 Data

Large amount of available data was used for verification of the theoretical assumptions. Those data are mostly spy game video, pictures, scout camera records, direct observations, animal footprints and others. The number of feral animals who emerged was registered at every underpass. Main task was to determine the number of feral animals who tried to go through the underpass. Subsequently the number of successful passes was evaluated. There are three categories of feral animals:

1. Large mammals (deer, bear, wolf, elk);
2. Medium mammals (roe-deer);
3. Small mammals (fox, marten, polecat and badger).

Large amount of data was obtained at a section of D1 highway Lipník nad Bečvou-Bělotín (former designation D4704). In this section there was installed camera system which monitored both sides of motorway and detailed record beneath the bridge. This allowed us to evaluate whether the species tried to go through and whether was this try successful. We obtained further data from highway sections D1, D11 and R35. We also consulted our uncertainties with animal behavior experts.

Table 1 contains data of 16 selected underpasses. Table shows the height, length and width of the underpasses in meters. It also shows number of animals which tried to go through (column total), number of successful passes (column 1), number of unsuccessful passes (column 0). The last column p indicates the ratio of successful passes to all passes. We assume that the probability of passing through will increase with increasing width and height and it will decrease with increasing length. Furthermore, we assume that the height will have more significant impact to increase the probability of migration than the width.

3 Methods

Observed dependent variable has only two possible outcomes. Animal goes through or it does not. Statistically we understand it as success or failure. We code this 1 or 0. Logit model was chosen as the best one for this task. Logit model is a special case of generalized linear model (GML). It is assumed that distribution of explained variable Y is alternative with the probability of succes π (probability of

Num.	Width	Length	Height	0	1	Total	p
1	50.80	40.90	7.70	10	412	422	0.976
2	177.60	25.50	22.94	0	56	56	1.000
3	62.70	25.50	8.33	1	36	37	0.973
4	789.50	25.50	13.14	0	412	412	1.000
5	83.00	25.00	7.07	9	248	257	0.965
6	70.80	27.40	6.98	25	689	714	0.965
7	10.00	37.50	7.14	74	341	415	0.822
8	18.00	25.50	9.51	10	623	633	0.984
9	18.00	34.20	11.52	1	126	127	0.992
10	11.60	38.00	6.05	109	545	654	0.833
11	12.00	39.90	6.07	41	218	259	0.842
12	11.30	33.40	5.07	80	277	357	0.776
13	11.30	33.00	5.08	82	287	369	0.778
14	9.00	36.50	4.20	70	88	158	0.557
15	13.43	25.50	3.25	104	261	365	0.715
16	17.25	25.50	5.44	79	710	789	0.900

Table 1 Analyzed data

animal going through in our case). We denote explaining variables as $X_1, \ldots, X_n$. The link function in logit model is in form:

$$g(\pi) = \log \frac{\pi}{1-\pi}, \tag{1}$$

where π represents the mean value of the random variable Y. Predictor of link function is linear combination of explaining variables

$$\alpha + \beta_1 X_1 + \cdots + \beta_n X_n, \tag{2}$$

where $\alpha, \beta_1, \ldots, \beta_n$ are unknown parameters. Unknown parameters are estimated by maximum likelihood method. It is possible to test the significance of each variable (factors) in the logit model. We can express estimation for the mean value of j-th random variable (estimation of technical migration potencial for j-th underpass) in this form:

$$\hat{\pi}_j = \frac{e^{\mathbf{X}^j \hat{\boldsymbol{\beta}}}}{1 + e^{\mathbf{X}^j \hat{\boldsymbol{\beta}}}}, \tag{3}$$

where $\mathbf{X}^j$ is vector $(1, X_{1j}, \ldots, X_{nj})$ and $\hat{\boldsymbol{\beta}}$ is vector of estimated unknown parameters. Interpretation of parameters in logit model is more difficult than in the linear one. Logit model says that if the estimation of the β_j parameter is positive then for increasing values of variable X_j with solid (not changing) values of other variables the probability of success will also grow. Negative values of β_j parameter will reduce the probability of success. Values of the parameter on its own cannot be interpreted. Value e^{β_j} can be interpreted as odds ratio which says how many times the probability of success will increase if the variable X_j is changing and other variables are constant.

At first we will use the logit model on all migration objects together. Later we will divide objects into "narrow" and "broad" category. Broad objects are numbers 1-6 in Table 1. The width (greater then 20 m) of these objects does not practically limit migrations. Narrow objects (width below 20 m) are numbers 7-16 in Table 1. It was used listed parameters and claster analysis for the separation.

4 Results

For data from Table 1 we have created logit model. To estimate unknown parameters we have used statistic software Gretl version 9.1.14. Results are shown in following table:

From values showed in Table 2 we can see that all quantities have influence (are statistically significant at the 5% significance level) on migration. As we assumed, length has negative influence on migration.

Parameter	Estimation	p-value
Constant	0.504	0.0618
Width	0.018	<0.0001
Length	-0.058	<0.0001
Height	0.492	<0.0001
Number of correctly estimated cases = 5329 (88.5 %)		

Table 2 Results of logit model

By increasing length the probability of go through is decreasing. More specifically every meter of extra length changes the probability 0.94 times. Width and height have positive influence; reality supports this because increasing these parameters will also increase the probability of passes. Every meter of width will increase the probability of pass 1.02 times. Every meter of height will increase the probability of pass 1.64 times. Results show that height has more significant impact on migration. This is expected conclusion because meter of height is optically more dominant then meter of width. Model correctly estimates 88.5 per cent of cases.

Later we dealt only with narrow underpasses (Table 1, objects 7-16). We created logit model again and got following results:

Parameter	Estimation	p-value
Constant	-2.654	<0.0001
Width	0.176	<0.0001
Height	0.357	<0.0001
Number of correctly estimated cases = 3476 (84.2 %)		

Table 3 Results of logit model for narrow underpasses

This model does not contain length quantity because we found it statistically insignificant at the 5% significance level. Width and height has positive influence on animals' pass. Increase of width and height also increases the probability of pass. More specifically every meter of width increase the probability 1.19 times. Every meter of height increase the probability 1.43 times. By comparing values from tables 2 and 3 we can see that width has greater influence at narrow underpasses. Probability increase of width used to be 1.02 now it is 1.19. Contrary, probability increase of height used to be 1.64 now it is only 1.43. Height is still more significant overall.

For broad bridges (Table 1, objects 1-6) we obtained following results:

Parameter	Estimation	p-value
Constant	-1.742	0.4238
Height	0.719	0.0174
Number of correctly estimated cases = 1853 (97.6 %)		

Table 4 Results of logit model for broad underpasses

This model contains only height parameter which is the only one statistically significant at the 5% significance level. As expected height has positive influence on go through probability. Increase of meter height also increases the probability 2.05 times. This model correctly estimates 97.6 per cent of all cases.

5 Enomomic analysis

Results of analysis by logit model are good standing point for economic analysis. Highway of type D27.5 and D33.5 and R24.5 are the most common ones in the Czech Republic. Length of the underpass is in the range between 26 and 39 meters. Length (following the methodology published in [15]) affects

technical migration potential only slightly. Logit model also shows small significance. Length is technically impossible to decrease due to width of the communication. We are not going to deal with it at all.

Designers usually have very limited option to change height of the underpass because the vertical alignment is given by much more significant factors. It conflicts with results that increase of height has positive influence on technical migration potential.

Width of the bridge can be changed easily by designers. It also depends on investor and his financial situation. This makes width important parameter which can increase highway permeability but also increase cost of the project.

Economic analysis takes prices of bridges in 2013 for highway type D27.5. For span under 12 meters reinforced concrete slabs were assumed. For span over 12 meters pre-tensioned girders with cast in place concrete slab were assumed. A parametric study of the motorway bridge with span between 4 and 30 meters was chosen as a practical example. Calculation was made for category of medium mammals (roe-deer) and for constant height of 3.5 meters. It is obvious from Table 5 that increasing span of the bridge (width of the underpass) increases significantly technical migration potential up to 15 meters. Further increasing of width of underpass does not come with increase of technical migration potential. Also span over 15 meters increases the cost significantly.

Span (m)	**Price (mil EUR)**	**Migration potential**
4	0.18	0.14
6	0.26	0.30
8	0.41	0.44
10	0.52	0.54
15	0.76	0.68
20	1.13	0.72
25	1.63	0.73
30	2.05	0.74

Table 5 Data for economic analysis

6 Conslusions

Analysis by logit model brought expected results. Increase of width of the underpass increases the migration potential value. More significant increase of migration potential comes along with increasing of the height. The length of the underpass is not significant. Increase of length slightly decreases the migration potential.

In narrow underpass cases the length is statistically insignificant parameter. Influence of width and height stays positive. Width has bigger influence than height. This matches with our expectations. In broad underpass cases, except length also width is insignificant which is understandable. Height is the only significant parameter. As expected, its increase has positive influence on migration potential.

Above mentioned results can be used for the economic design of the measures to enable the permeability of line structures. The utility of the existing as well as the newly designed bridge objects should be increased instead of building new expensive underpasses and overpasses.

Acknowledgements

The support of Eureka project E! 6752 DETECTGAME is appreciated.

References

[1] Ament, R., Clevenger, A. P., Wu, O. and Hardy, A.: An assessment of road impacts on wildlife populations in U.S. National Parks, *Environmental Management* **42** (2008), 480–496.

[2] Austin, J. M., and Garland, L.: Evaluation of a wildlife underpass on Vermont State Highway 289 in Essex, Vermont. In: *Proceedings of the 2001 International Conference on Ecology and Transportation* (C. L. Irwin, P. Garrett, K. P. McDermott, eds.), Center for Transportation and the Environment, North Carolina State University, Raleigh, 2001, 616–624.

[3] Clevenger, A. P., Huijser, M. P.: *Wildlife Crossing Structure Handbook, Design and Evaluation in North America*. Federal Highway Administration, Washington D.C., USA, 2011.

[4] Clevenger, A. P.: 15 Years of Banff Research: What We Have Learned and Why It is Important to transportation Managers beyond the Park Boundary. In: *Proceedings of the 2011 International Conference on Ecology and Transportation*, 2012, 433–447.

[5] Crooks, K., Haas, C., Baruch-Mordo, S., Middledorf, K., Magle, S., Shenk, T., Wilson, K., and Theobald, D.: *Roads and connectivity in Colorado: animal-vehicle collisions, wildlife mitigation structures, and lynx-roadway interactions*. Colorado Department of Transportation, Denver, Colorado, 2008.

[6] Crooks, K. R., and Sanjayan, M.: *Connectivity conservation*. Cambridge University Press, Cambridge, UK, 2006.

[7] Davenport, J. and Davenport, J. L.: *The ecology of transportation: managing mobility for the environment*. Springer, London, UK, 2006.

[8] Johnson, B. L., Hildreth, A. M., Hygnstrom, S. E., VerCauteren, K. C.: Creating resilient wildlife undercrossings with interlocking pavement blocks. In: *Proceedings of the 2013 International Conference on Ecology and Transportation*. Retrieved from http://www.icoet.net/ICOET_2013/proceedings.asp.

[9] Hilty, J., Lidicker, W., and Merenlender, A.: *Corridor ecology: the science and practice of linking landscapes for biodiversity conservation*. Island Press, Washington, D.C., 2006.

[10] Huijser, M. P., Duffield, J. W., Clevenger, A. P., Ament, R. J., and McGowen, P. T.: Cost-benefit analyses of mitigation measures aimed at reducing collisions with large ungulates in the United States and Canada: a decision support tool, *Ecology and Society* **14** (2009), 15.

[11] Kendall, K. C., Stetz, J. B., Roon, D. A., Waits, L. P., Boulanger, J. B., and Paetkau, D.: Grizzly bear density in Glacier National Park, Montana, *Journal of Wildlife Management* **72** (2008), 1693–1705.

[12] Little, S. J., Harcourt, R. G., and Clevenger, A. P.: Do wildlife passages act as prey traps? *Biological Conservation* **107** (2002), 135–145.

[13] Mata, C., Hervás, I., Herranz, J., Suárez, F., and Malo, J. E.: Complementary use by vertebrates of crossing structures along a fenced Spanish motorway. *Biological Conservation* **124** (2005), 397–405.

[14] Millions, D. G., and Swanson, B. J.: Impact of natural and artificial barriers to dispersal on the population structure of bobcats. *Journal of Wildlife Management* **71** (2007), 96–102.

[15] Myšková, K., Žák, J.: Modelling of Wildlife Migrations and its Economic Impacts, *AIP Conference Proceedings* **1558** (2013), 1867–1870.

[16] Pearce, J. L., and Boyce, M. S.: Modelling distribution and abundance with presence-only data, *Journal of Applied Ecology* **43** (2006), 405–412.

[17] Rykiel, E. J.: Testing ecological models: The meaning of validation, *Ecological Modelling* **90** (1996), 229–244.

[18] Van der Grift, E. A., van der Ree, R., Fahrig, L., Findlay, S., Houlahan, J., Jaeger, J. A. G., Klar, N., Madriñan, L. F., Olson, L.: Evaluating the effectiveness of road mitigation measures, *Biodiversity and Conservation* **22** (2013), 425–448.

[19] Van der Grift, E. A., Schippers, P.: Wildlife crossing structures: can we predict effects on population persistence? In: *Proceedings of the 2013 International Conference on Ecology and Transportation*. Retrieved from http://www.icoet.net/ICOET_2013/proceedings.asp

[20] Žák, J., Florian, A.: Green Bridges and their Migration Potential, *International Journal of Civil, Architectural Science and Engineering* **7** (2013), 890–896.

The use of fuzzy logic in the operational risk assessment of mining companies

Tomasz Leszek Nawrocki[1], Izabela Jonek-Kowalska[2]

Abstract. The main objective of this paper is to present the concept and practical application of the operational risk assessment fuzzy model. This model is an attempt to combine fuzzy methodology with financial indicators. When creating the basic assumptions of the model, the authors used a resource approach, treating human, material and financial resources, and the way they are allocated, as the detailed sources of operational risk. In addition, the model also takes into account the relational and organizational resources, as well as dependence of operational risk on the surroundings. In this article, the model has been verified by using it to perform an operational risk assessment in the mining industry on the example of coal mining companies listed on the Warsaw Stock Exchange.

Keywords: operational risk, fuzzy logic, fuzzy model, coal mining companies**.**

JEL Classification: G32, C69
AMS Classification: 94D05, 26E50

1. Introduction

Identification and measurement of operational risk is a complex and multi-threaded issue in which the basic problem is to anticipate economic events and their influence on the financial results of a company. In the subject literature, the methodology of risk assessment is most often presented in the context of econometric-statistical or probabilistic methods [8]. However, their use frequently refers to only some parts of a company's activity, e.g. investment risk or market risk. There is still a lack of clear, reliable and holistic methods of risk assessment of the company's activity that are characterized by comparability and easy application by all the company's stakeholders. Therefore, the main purpose of this article is to present a comprehensive conception and practical application of the operational risk assessment model based on fuzzy set theory [5], [10].

This model is an attempt to combine the methodology of fuzzy sets with the ratios of financial analysis. The presented model has been positively assessed in the theoretical part by domestic reviewers [4]. Currently, the authors are searching for further approaches to model improvement in the course of empirical research; thus, in this article, the model was subject to verification by way of using it for operational risk assessment in the mining industry, specifically the selected coal companies listed on the Warsaw Stock Exchange. The results of the conducted research enable an individualized assessment of operational risk, industrial comparisons and further model improvement.

2. Research methodology

In the proposed solution, taking into account the essence of operating activity, it was assumed that the operational risk could be analysed and assessed in two dimensions: resource potential and the specific nature of the business run. In the assessment of resource potential, there are three basic groups of resources, understood in a broad sense, involved: human resources, tangible and intangible resources and financial resources. In turn, in cases of risk assessment stemming from the specific nature of a company's operating activity, the focus was on the risk factors determining the operating results of the company, the complexity of the company's activity and its relationships with suppliers and recipients.

The structure of the suggested operational risk assessment model, along with the most detailed assessment criteria within the particular modules, is presented in Figure 1.

[1] Silesian University of Technology, Faculty of Organization and Management, Institute of Economics and Informatics, Roosevelta 26 Str., 41-800 Zabrze, Poland, e-mail: tomasz.nawrocki@polsl.pl.
[2] Silesian University of Technology, Faculty of Organization and Management, Institute of Economics and Informatics, Roosevelta 26 Str., 41-800 Zabrze, Poland, e-mail: izabela.jonek-kowalska@polsl.pl.

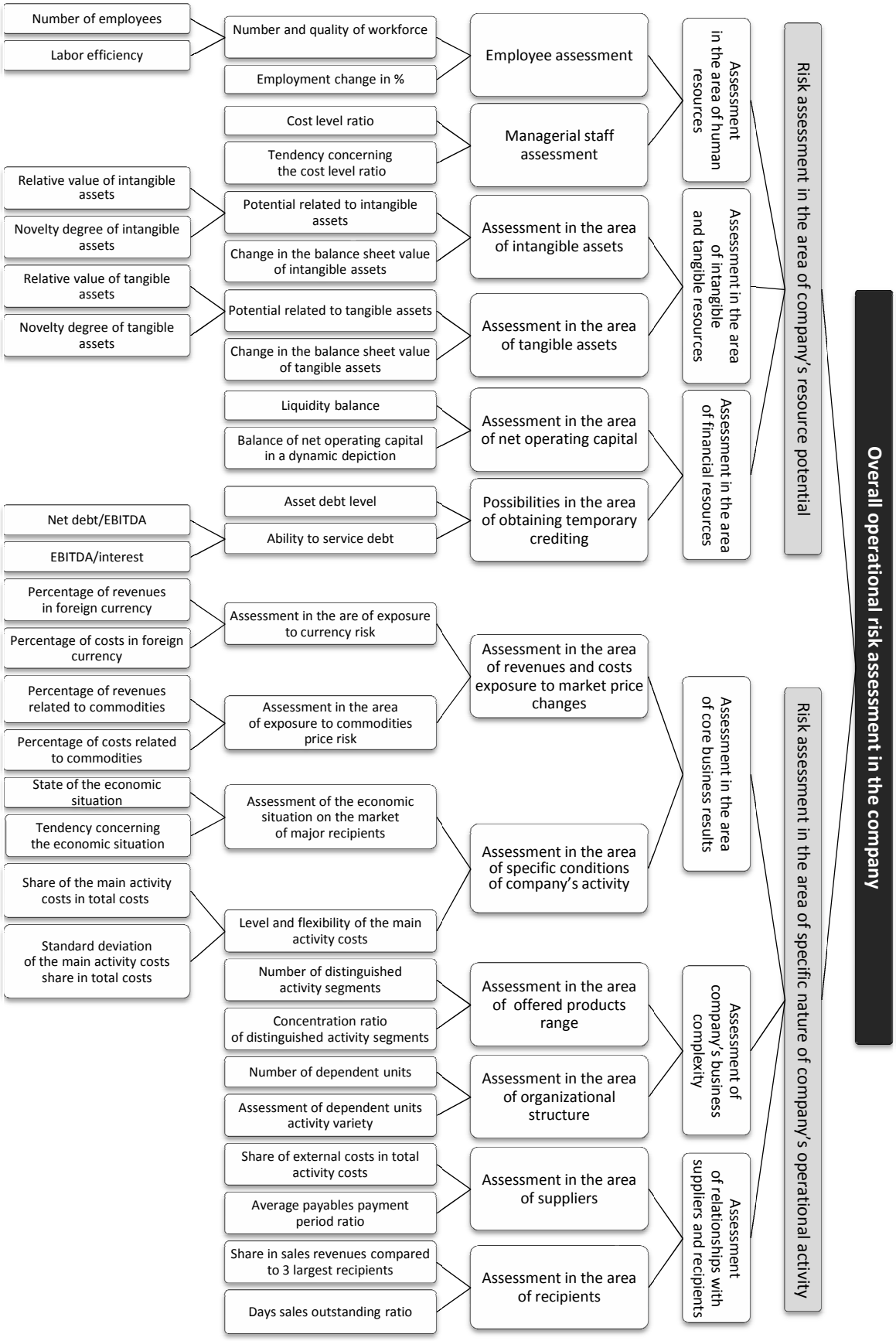


Figure 1 Structure of the operational risk assessment model
Source: Authors' work

In the proposed model firstly it is intended to obtain partial assessments within the distinguished basic assessment criteria of operational risk. These assessments will result from the ratios calculated on the basis of data from financial statements or, if no possibility exists to make such calculations, a qualitative (descriptive) assessment of a particular criterion stemming from the description or characteristics in the periodical report of the examined company. Next, on this basis, aggregated assessment results may be obtained in the areas of human resources, tangible and intangible resources, financial resources, core business results, company's business complexity and its relationships with suppliers and recipients. Furthermore, these results constitute foundations for calculating general risk measures in the areas of resource potential and the specific nature of the company's operational activity, so that in the final stage, on their basis, it is possible to achieve overall operational risk assessment for the analysed company.

The calculation aid in the suggested solution is based on the fuzzy set theory, which is one of the approximate reasoning methods [2], [3]. In classic set theory, the transition from the full membership of an element in a set to its total non-membership is bivalent (either the element is the member of a set or it is not), thus presenting imprecise concepts using such sets raising a number of issues. L.A. Zadeh, the author of fuzzy set theory, was the first one to notice that, when formulating the rule of inconsistency: „*complexity and precision bear an inverse relation to one another in the sense that, as the complexity of a problem increases, the possibility of analyzing it in precise terms diminishes*" [11]. In fuzzy set theory, it is accepted that the element may partially belong to a set and at the same time to its complement; therefore, the law of excluded middle does not apply here. In the fuzzy set, the transition from membership into a set to non-membership is gradual (this gradual change is expressed by the so-called membership function), thus these concept makes it possible to describe soft concepts and imprecise quantities, to which the assessment of the company's operational activity risk certainly belongs. From the formal point of view, above issues are expressed by the following definition of a fuzzy set [6]:

Definition 1. *A fuzzy set* A *in a certain space (area of consideration)* $X = \{x\}$ (*which is written as* $A \subseteq X$) *is a set of pairs*: $A = \{(\mu_A(x), x)\}$, $x \in X$, *where* $\mu_A(x)$: $X \rightarrow [0, 1]$ *is the membership function of fuzzy set* A, *to which each element* $x \in X$ *attributes its degree of membership in the fuzzy set* A, $\mu_A(x) \in [0, 1]$.

An important concept of fuzzy logic is the linguistic variable. Even though the mathematical formalism of this variable is relatively complicated, its intuitive meaning is simple – a linguistic variable is a variable, which values are not numbers but sentences in a certain language, identified in the semantic sense with particular fuzzy sets [7]. In turn, the basic mean enabling the presentation of the relations occurring between the accepted linguistic variables are fuzzy conditional sentences in the form:

IF *x* is *A* THEN *y* is *B*.

Usually, however, the relation between the same variables is described not by a single rule but by the so-called bank (base) of rules, which is treated in the process of fuzzy reasoning as a certain whole – a subsystem, the total effect of which is subjected to further processing. In the process of reasoning, for given inputs, all of the rules in the bank are activated, and the results of their actions are then merged into a fuzzy output set, which is the value of the variable *y*. The given bank of rules may describe the relations between the input and output of the entire system, or it can be an element of a more complex hierarchical structure [1]. The equivalent of the real system model in the fuzzy logic is a fuzzy-model. In literature, there are various types of fuzzy models characterized, but one of the most popular is the Mamdani model, which general diagram is presented in Figure 2.

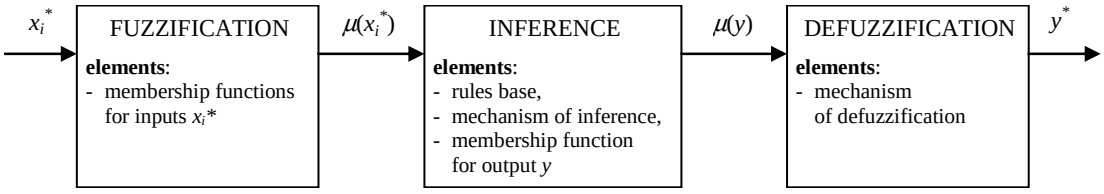


Figure 2 General diagram of Mamdani fuzzy model
Source: Authors' work, based on: Piegat A., Modelowanie i sterowanie rozmyte, EXIT, Warsaw 2003, p. 165

As the input of the fuzzy model x_i^* values are introduced, which in the FUZZIFICATION module are subjected to the fuzzification process – here, the membership degree of input values $\mu(x_i^*)$ to the particular fuzzy sets is calculated. Next, in the INFERENCE module, based on the received input membership degrees, fuzzy reasoning takes place, the end result of which is a resultant membership function $\mu(y)$ of the model output. The basis for the fuzzy reasoning is so-called rule base in the form of „IF – THEN" and an inference mechanism which determines the way of activating rules in the base, as a result of which membership functions

of the particular rules conclusions with the given values of fuzzy model inputs x_i^* are received, as well as the way for their aggregation into one resultant membership function of the entire base conclusion $\mu(y)$. Because this function most often has a fuzzy form, which makes the interpretation of the final result much more difficult, in many cases, there is a need to transform it into a precise value. This is done in the DEFUZZIFICATION module, where by the use of the chosen method a sharp (non-fuzzy) value of the model output y^* is calculated [9].

3. Detailed assumptions of the research

In order to verify the proposed model the operational risk assessment was conducted for two mining enterprises, organized in the form of capital groups, which shares are listed on the Warsaw Stock Exchange – Jastrzębska Spółka Węglowa S.A. (JSW) and Lubelski Węgiel „Bogdanka" S.A. (LW Bogdanka).

According to the adopted methodology, the basis for the operational risk assessment of mentioned above entities were the data acquired from the consolidated periodical reports (annual, semi-annual, quarterly) and other materials (presentations of results, marketing reports) published by these companies in the years 2011-2014. In relation to the construction of the operational risk assessment fuzzy model, based on Mamdani approach, the following assumptions were made:

- for all input variables of the model, the same dictionary of linguistic values was used, and their space was divided into three fuzzy sets, most often named {low, medium, high};
- for output variables of the model, the space of linguistic values was divided into five fuzzy sets named {low, mid-low, medium, mid-high, high};
- in the case of all membership functions to the particular fuzzy sets, a triangular shape was decided for them (Figure 3 and Figure 4);
- the values of the characteristic points of fuzzy sets (x_1, x_2, x_3) for the particular input variables of the model were determined arbitrarily, based on the distribution of analysed variables values and on the authors many years' experience in the area of financial and risk analysis;
- for the fuzzification of the input variables, the method of *simple linear interpolation* was used [1];
- fuzzy reasoning in the particular knowledge bases of the model was conducted using *PROD* operator (fuzzy implication) and *SUM* operator (final accumulation of the conclusion functions received within the particular rule bases into one output set for each base) [9];
- for the defuzzification of fuzzy reasoning results within the particular rule bases *Center of Sums* method was used [11].

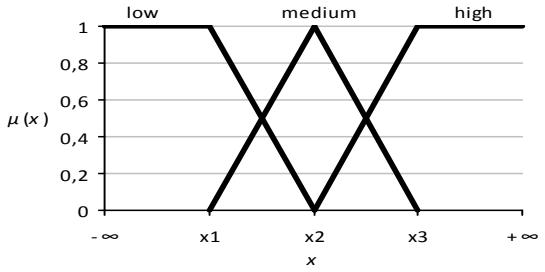


Figure 3 The general form of the input variables membership function to distinguished fuzzy sets. Source: Authors' work

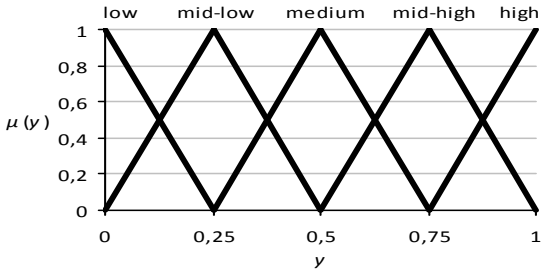


Figure 4 The output variables membership function to distinguished fuzzy sets. Source: Authors' work

Next, taking into consideration the general structure of the operational risk assessment model presented in Figure 1, the authors, based on their knowledge and experience in the area of the analysed issue, designed 29 bases or rules in the form of „IF – THEN" (27 bases with 9 rules and 2 bases with 27 rules), achieving this way a „ready to use" form of the operational risk assessment fuzzy model. The intermediate and final assessments generated by the model take values in the range between 0 and 1, where from the viewpoint of the analysed issue, the values closer to 1 mean a very favourable result (lower risk), while the values closer to 0 indicate a result less favourable (higher risk).

All calculations related to the presented fuzzy model were based on self-developed structure of formulas in MS Excel.

4. Research results

As emphasized in the methodological section, the final operational risk assessment in the examined mining companies consists of the results in the two basic areas: resources determining the operating activity,

and its specific conditions. The assessment results of the first one are presented in Figure 5. In the whole analysed period, a lower operational risk in the area of resources (expressed by a higher assessment value) characterizes LW Bogdanka. This enterprise is distinguished by the high and stable assessment results of human resources stemming from the highest labour efficiency in the Polish mining industry and very good cost level ratio. Nevertheless, it should be added that these two parameters are dependent in a great part on favourable geological-mining conditions. LW Bogdanka is characterised by a very low intensity of natural hazards and high thickness of deposits localized relatively shallowly. In this company, since the beginning of 2013, the situation concerning the assessment of financial resources has deteriorated, mainly due to a lack of compatibility of the need for net operating capital with its real estate. The assessment results in the area of tangible and intangible resources have decreased as well. In the second of the examined enterprises (JSW), the level of operational risk in the area of resources was clearly increasing in the year 2013 (lower assessment value). However, the company, over the whole analysed period, maintains a high and stable assessment result in the area of financial resources, what proves an effective financial management oriented to limitation of the economic fluctuation influence on financial conditions. Nonetheless, the assessment results in the area of human resources have been decreasing with time, connected with efficiency deterioration as a result of production reduction. In 2013, import of coking coal to Poland increased, mostly from Australia and the Czech Republic. The company is also under a strong influence of economic fluctuations on the European steel market. In the year 2013 the assessment results of tangible and intangible resources deteriorated in this company too.

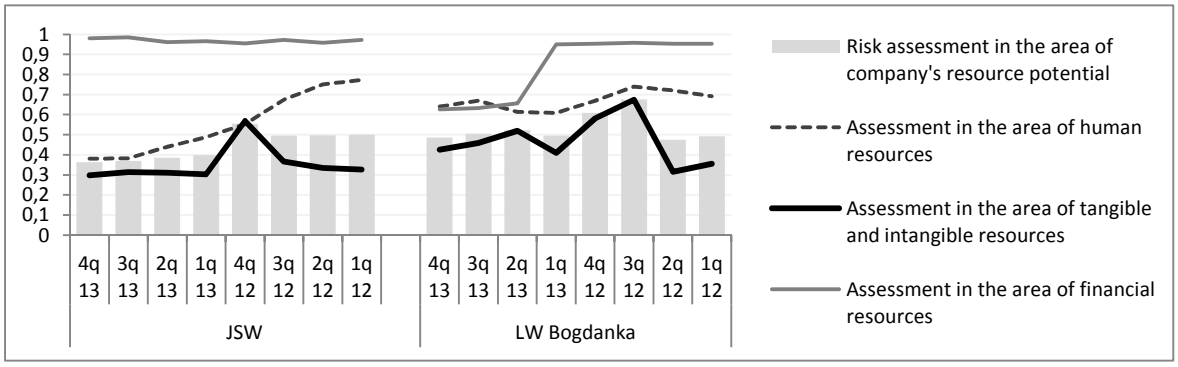


Figure 5 Risk assessment in the area of resource potential in the examined companies
Source: Authors' work

Assessment in the area of the specific nature of operating activity is quite stable over time in both companies and clearly lower in JSW. The latter mostly stems from a greater exposure of the company to the risk of market price changes and economic fluctuations on the market of coking coal, mainly driven by economic fluctuations on the market of this raw resource recipients. Risk in this area is also enhanced by the low cost flexibility and low assessment results of organizational structure, coming from larger complexity and activity diversification of dependent units. LW Bogdanka characterizes lower risk in the area of operating activity specific nature, mostly due to a lower exposure to currency and price risk and slight economic changes on the market of major recipients who are, in this case, the local energy producers in the sector of industrial energy.

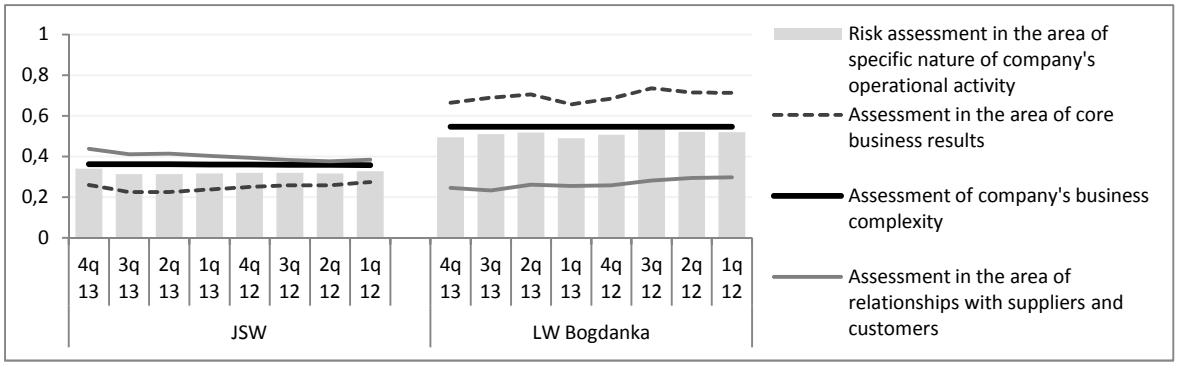


Figure 6 Risk assessment in the area of the specific nature of operational activity in the examined companies
Source: Authors' work

A resultant operational risk assessment presented in Figure 7 is an effect of assessment results in the aforementioned partial areas. Its variability in time is therefore shaped in great part by variability in the area of resources potential. The assessment of the company's activity specific nature, due to stability over time, only affects the level of the final assessment. In terms of results, operational risk is higher in the case

of JSW, the company providing resources for more demanding and geographically dispersed recipients. LW Bogdanka is characterized by highly assessed resource potential and lower exposure to market, price and currency fluctuations, which affect the lower level of operational risk.

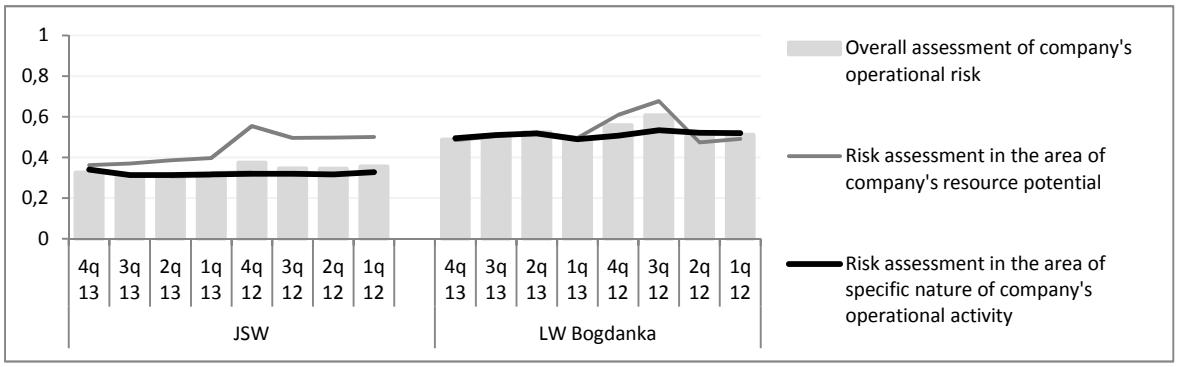


Figure 7 Overall assessment of operational risk in the examined companies
Source: Authors' work

5. Conclusions

The model of operational risk assessment, elaborated using the methodology of fuzzy sets, is characterized by the following advantages that partially enable reducing the defects of methods previously applied in risk identification and assessment in the company: combination of analytic and synthetic operational risk assessment, achieved through the use of sub-criteria that are later aggregated into a collective risk assessment; combination of quantitative and qualitative approach to risk assessment, manifested in a qualitative dimension of ratios used for risk measurement and qualitative method of risk diversification determination; possibility of conducting assessment by the external and internal stakeholders without the necessity of using the advanced calculation techniques; using for the assessment process data included in the generally accessible business periodical reports; providing reliability and flexibility at the same time due to the use of quantitative and qualitative data.

Operational risk assessment conducted using presented model allows the formulation of the following practical conclusions: operational risk in the whole examined period is higher in JSW than in LW Bogdanka, which is mostly caused by the lower results of risk assessment in JSW in the area of resource potential and higher exposure of this company to market, price and currency risk, as well as by the larger geographical dispersion of recipients; in both examined companies, operational risk clearly increased in the year 2013 in connection with the inflow of cheaper power and coking coal from import; risk assessments in the area of specific nature of company's operational activity in the examined entities are much more stable over time than risk assessments in the area of resource potential.

References

[1] Bartkiewicz, W.: Zbiory rozmyte. In: *Inteligentne systemy w zarządzaniu*. (Zieliński, J.S., ed.). PWN, Warszawa, 2000, 72-140.
[2] Farhadinia, B.: A series of score functions for hesitant fuzzy sets, *Information Sciences* 277 (2014), 102-110.
[3] Hejazi, R., S., Doostparast, A. Hosseini, S.M.:An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers, *Expert Systems with Applications* 38 (2011), 9179-9185.
[4] Jonek-Kowalska, I., and Nawrocki, T.: Koncepcja rozmytego modelu oceny ryzyka operacyjnego w przedsiębiorstwie. In: *Finansowe uwarunkowania rozwoju organizacji gospodarczych. Ryzyko w rachunkowości i zarządzaniu finansami* (Turyna, J., Rak, J., eds.). Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego, Warszawa, 2013, 539-559.
[5] Kumar, G., Maiti, J.: Modeling risk based maintenance using fuzzy analytic network process, *Expert Systems with Application*s 39 (2012), 9946-9954.
[6] Kacprzyk, J.: *Wieloetapowe wnioskowanie rozmyte*, WNT, Warszawa, 1986.
[7] Łachwa, A.: *Rozmyty świat zbiorów, liczb, relacji, faktów, reguł i decyzji*, EXIT, Warszawa, 2001.
[8] Mandal, S., Maiti, J.: Risk analysis using FMEA: Fuzzy similarity value and possibility theory based approach, *Expert Systems with Applications* 41 (2014), 3527-3537.
[9] Piegat, A.: *Modelowanie i sterowanie rozmyte*, EXIT, Warszawa, 2003.
[10] Wulan, M., Petrovic, D.: A fuzzy logic based system for risk analysis and evaluation within enterprise collaborations, *Computers in Industry* 63 (2012), 739-748.
[11] Zadeh, L.A.: Fuzzy sets, *Information and Control* 8 (1965), 338-353.

Efficiency of the European labour markets: The Case of Czech Republic (A Stochastic frontier model approach)

Daniel Němec[1]

Abstract. This contribution aims to provide a consistent methodology to evaluate the performance of the European labour markets in the last 20 years and to reveal the most important factors influencing the efficiency of these labour markets. In particular, it focuses on the Czech labour market. Labour markets and their dynamics may be described by the standard Cobb-Douglas matching functions. Successful matches are thus treated as an output of a production process where unemployed are paired with available vacancies. Unemployment outflows are determined by the efficiency of corresponding matching process. Using stochastic frontier model approach, we estimate the efficiency of matching functions of Czech labour market, reveal the dynamics of quantified efficiency indicators and evaluate the differences among the regions. The stochastic frontier is estimated using the regional panel data set for the period 1997-2013. The model specification includes fixed effect term where individual effect terms and inefficiency terms are estimated jointly.

Keywords: matching efficiency matching function, European labour markets, Czech Republic, stochastic frontier model, panel data.

JEL Classification: J41, C23, E24
AMS Classification: 62P20

1 Introduction

Labour market dynamics is influenced by many economic and institutional factors. The institutional factors may be connected with labour market efficiency. Most of the concepts dealing with the "efficiency" approach are based on the matching function framework. In this framework, successful labour market matches are treated as an outcome of interactions between unemployed job seeker and vacancies. Quantifying the efficiency of the European labour markets belongs to the highly relevant question in the last decade. Unfortunately, using the aggregate labour market statistics for European countries does not allow us to estimate directly the corresponding efficiencies. We are able to obtain the time series of vacancies and unemployed but the successful matches are missing and cannot be recovered in an easy way. This kind of statistics must be obtained from the regional labour market statistics for each country. Panel data structure is thus rich enough to use stochastic frontiers model approach separately for country specific regional labour markets.

This contribution aims to provide a consistent methodology to evaluate and quantify the effectiveness of the Czech labour market from the view of regional labour markets. We are using the stochastic frontier panel data model approach with quarterly regional data and explicitly treated fixed effects term in the matching function model equation. On the one hand, this approach extends the previous investigations of the efficiency of the Czech labour market carried out by Němec [5], [6] or by Tvrdoň and Verner [7]. Their results have been based on the aggregate labour market statistics. On the other hand, using the data from quarterly regional labour market statistics and stochastic frontier panel data model methodology, it offers a new insight into the outcomes of the Czech labour market in the last 15 years and extends the detailed analysis of Galuščák and Münich [2] in a specific way, i.e. by dealing with efficiency issues. Stochastic frontier model approach has been used by Ilmakunnas and Pesola [4] in their study of regional labour markets in Finland. They used annual data and did not take into account explicitly possible individual fixed effects of examined regions. Gorter et al. [3] investigated the efficiency in the Dutch labour market in Netherland along the same lines. They have observed that the estimated labour market efficiency increases during the recession and recovery period while it decreases during the economic booms. This interesting feature is considered in this contribution as well.

2 Stochastic frontier model with panel data

Stochastic frontier model approach is a modern econometric tool that allows us to measure the performance of production units. Production technology is described by the production function where inputs are transformed

[1] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, nemecd@econ.muni.cz

into outputs of any kind. This parametric approach to measure technical inefficiency is very flexible and may be used in many applications. As for the labour market framework, the production technology of a labour market is usually described by the matching function.

2.1 Matching function and matching efficiency

Dynamic relationship between the flows of unemployed and the flows of unfilled job vacancies can be described simply by a standard production function with two inputs: the unemployed and the vacancies. New matches are thus an outcome of this matching process. In this contribution, the regional labour markets are represented by a standard Cobb-Douglas matching function in log-linear form:

$$\ln M_{it} = \alpha_i + \beta_{\log(u)} \ln U_{it} + \beta_{\log(v)} \ln V_{it} + \varepsilon_{it}, \tag{1}$$

where $i = 1,\ldots,N$ denotes the regions and $t = 1,\ldots,T$ the time periods. The α_i terms are fixed regional effects and ε_{it} represents stochastic factors which will be discussed below. This basic form of matching function may be extended and modified in many ways. Ilmakunnas and Pesola [4] implemented regional and labour force characteristics directly into the matching function by means of other explanatory variables. Resulting efficiency was thus a linear function of regional fixed effects and various regional characteristics. In their view, the term ε_{it} was treated purely as white noise process. Similar approach may be found in the work of Gorter et al. [3]. Galuščák and Münich [2] enhanced the basic matching function form by the flow factors (i.e. unemployment and vacancy inflows realized during the time period). Stochastic frontier model approach tries to model the stochastic term ε_{it} as consisting of combination of random variations in the matching process and the region specific inefficiency term. Regional and labour force characteristics are then implemented directly into this inefficiency term. This approach was used by Ilmakunnas and Pesola [4]. But they did not include the fixed (or random) region effects. In this contribution, we try to estimate the inefficiency of the Czech regional labour market using fixed effect panel stochastic model. This model approach is thus able to capture region specific individual effects, basic matching function characteristics and time-varying regional inefficiency terms at once.

2.2 Fixed effect panel stochastic model

To estimate matching function parameters and the inefficiency of the matching process we use the approach proposed by Wang and Ho [8]. Their specification of a stochastic frontier model is as follows:

$$y_{it} = \alpha_i + \mathbf{x_{it}}\boldsymbol{\beta} + \varepsilon_{it}, \tag{2}$$

$$\varepsilon_{it} = v_{it} - u_{it}, \tag{3}$$

$$v_{it} \sim N\left(0, \sigma_v^2\right), \tag{4}$$

$$u_{it} = h_{it} \cdot u_i^*, \tag{5}$$

$$h_{it} = f\left(\mathbf{z_{it}}\boldsymbol{\delta}\right), \tag{6}$$

$$u_i^* \sim N^+\left(\mu, \sigma_u^2\right), \qquad i = 1,\ldots,N \qquad t = 1,\ldots,T. \tag{7}$$

In this model framework, α_i is individual fixed effect for the unit i, $\mathbf{x_{it}}$ is a $1\times K$ vector of explanatory variables, v_{it} is a random error with zero mean, u_{it} is a stochastic variable measuring inefficiency, and h_{it} is a positive function of a $1\times L$ vector of non-stochastic determinants of inefficiency ($\mathbf{z_{it}}$). Constant term is excluded from explanatory variables and inefficiency determinants. It should be clear that the notation N^+ means that the realized values of the variable u_i^* are positive. In case of $\mu = 0$ the variable u_i^* follows a half-normal distribution. Wang and Ho [8] showed how to remove the fixed individual effect from the model. This procedure allows us to estimate all the model parameters. Of course, the individual effects may be recovered from the final parameter estimates. There are two possible approaches to model transformation: first-differencing and within-transformation. Both methods are equivalent (see Wang and Ho [8]). Stochastic frontier model of the Czech regional labour markets has been identified using the first-difference transformation. The main points of this methodology may be described as follows (for detailed discussion see Wang and Ho [8]).

It is necessary to define first difference of corresponding variables as $\Delta w_{it} = w_{it} - w_{it-1}$ and the stacked vector of Δw_{it} for a given i and $t = 2,\ldots,T$ is denoted as $\Delta\tilde{w}_i = \left(\Delta w_{i2}, \Delta w_{i3}, \ldots, \Delta w_{iT}\right)'$. Assuming that the function h_{it} is not constant, i.e. the vector z_{it} contains at least one time-varying variable, the model in its first-difference form may be expressed as:

$$\Delta\tilde{y}_{it} = \Delta\tilde{\mathbf{x}}_{\mathbf{it}}\boldsymbol{\beta} + \Delta\tilde{\varepsilon}_i, \tag{8}$$

$$\Delta\tilde{\varepsilon}_i = \Delta\tilde{v}_i - \Delta\tilde{u}_i, \tag{9}$$

$$\Delta\tilde{v}_i \sim MN(0,\Sigma), \tag{10}$$

$$\Delta\tilde{u}_i = \Delta\tilde{h}_i u_i^*, \tag{11}$$

$$u_i^* \sim N^+\left(\mu,\sigma_u^2\right), \qquad i = 1,\ldots,N \tag{12}$$

It is obvious from panel data models that first-difference introduces correlations of Δv_{it} within the *i*th panel. The covariance matrix of the multivariate distribution of $\Delta\tilde{v}_i$ is

$$\Sigma = \begin{bmatrix} 2\sigma_v^2 & -\sigma_v^2 & 0 & \cdots & 0 \\ -\sigma_v^2 & 2\sigma_v^2 & -\sigma_v^2 & \cdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\sigma_v^2 \\ 0 & 0 & \cdots & -\sigma_v^2 & 2\sigma_v^2 \end{bmatrix}. \tag{13}$$

The covariance matrix Σ has elements $2\sigma_v^2$ on the diagonal and $-\sigma_v^2$ on the off-diagonal. The key point revealed by Wang and Ho [8] is that the distribution of the term u_i^* is unaffected by the transformation. This fact helps to derive marginal log-likelihood function for each panel unit:

$$\ln L_i = -\frac{1}{2}(T-1)\ln(2\pi) - \frac{1}{2}(T-1)\ln\left(\sigma_v^2\right) - \frac{1}{2}\Delta\tilde{\varepsilon}_i'\Sigma^{-1}\Delta\tilde{\varepsilon}_i + \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right) + \ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right) - \ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right) \tag{14}$$

where

$$\mu_* = \frac{\mu/\sigma_u^2 - \Delta\tilde{\varepsilon}_i'\Sigma^{-1}\Delta\tilde{h}_i}{\Delta\tilde{h}_i'\Sigma^{-1}\Delta\tilde{h}_i + 1/\sigma_u^2} \qquad \sigma_*^2 = \frac{1}{\Delta\tilde{h}_i'\Sigma^{-1}\Delta\tilde{h}_i + 1/\sigma_u^2} \qquad \Delta\tilde{\varepsilon}_i = \Delta\tilde{y}_i - \Delta\tilde{\mathbf{x}}_{\mathbf{i}}\boldsymbol{\beta}. \tag{15}$$

In this expression, Φ is the cumulative density function of a standard normal distribution. Log-likelihood function of the model is obtained by summing the above function over al panel units.

Wang and Ho [8] approximated the observation specific technical inefficiency as conditional expectation

$$E\left(u_{it}\middle|\Delta\tilde{\varepsilon}_i\right) = h_{it}\left[\mu_* + \frac{\phi(\mu_*/\sigma_*)\sigma_*}{\Phi(\mu_*/\sigma_*)}\right] \tag{16}$$

evaluated at estimated values of term $\Delta\tilde{\varepsilon}_i$. This is a modified estimator of inefficiency terms which uses $\Delta\tilde{\varepsilon}_i$ instead of $\tilde{\varepsilon}_i$ as the conditional term. The main advantage is that the vector $\Delta\tilde{\varepsilon}_i$ contains all the information of individual unit in the sample and does not depend on individual effect term α_i that has the variance of higher order in case of small time dimension of the sample (variance of order $1/T$ in comparison to the variance of $1/((N-1)T)$ for the estimator $\hat{\beta}$). Technical efficiency may be obtained in accordance with other studies (see Battese and Coelli [1]) as $\exp(-u_{it})$. For derivation of individual fixed effects terms see Wang and Ho [8].

2.3 Data and model specification

The model for the Czech regional labour markets is estimated using the quarterly data set covering a sample of 14 regions from the 2nd quarter 1997 to the 2nd quarter 2013. The original data come from database of the Ministry of Labour and Social Affairs which cover the monthly data from regional Employment offices (77 districts). I used the following variables: the number of registered successful matches, M_{it}, in the corresponding month, the number of unemployed at the start of the month, U_{it}, and the number of vacancies at the start of the month, V_{it}. All the data are seasonally unadjusted. The quarterly data were transformed as monthly averages and summed up across the districts corresponding to the specific region.

Panel data set consists of 14 regions and 65 quarters. The estimated model has the form defined by the equations (2)-(7), where $y_{it} = \ln M_{it}$, $\mathbf{x}_{\mathbf{it}} = (\ln U_{it}, \ln V_{it})$, $\boldsymbol{\beta} = \left(\beta_{\log(u)}, \beta_{\log(v)}\right)'$, and

$$h_{it} = \left|\delta_{quarter}q_t + \delta_{quarter^2}q_t^2 + \delta_{quarter^3}q_t^3 + \delta_{\Delta\log GDP}\Delta\log GDP_t + \delta_{\Delta\log INV}\Delta\log INV_t\right|.$$

The variable q represents the time trend in a year, i.e. $q = \{1,2,3,4\}$. Variables $\Delta \log GDP_t$ and $\Delta \log INV_t$ are the quarter-on-quarter growths of gross domestic product and gross fixed capital formation in the Czech Republic respectively. These statistics are obtained from quarterly national accounts provided by the Czech statistical office. We have assumed $\mu = 0$ and we have thus a half-normal representation of the model. Inefficiency terms capture time trend (which is usual in many applications; see e.g. Battese and Coelli [1]). For computational purposes, the variance parameters were parameterised as $c_v = \log \sigma_v^2$ and $c_u = \log \sigma_u^2$ respectively.

3 Efficiency estimates

The model parameters were estimated by numerically maximizing the sum of marginal log-likelihood functions (14). All the estimation procedures were performed using Matlab version 2013b and its implemented function for unconstrained maximization.

$\beta_{\log(u)}$	$\beta_{\log(v)}$	$\delta_{quarter}$	$\delta_{quarter^2}$	$\delta_{quarter^3}$	$\delta_{\Delta \log GDP}$	$\delta_{\Delta \log INV}$	$\log \sigma_v^2$	$\log \sigma_u^2$
0.470	0.058	-0.551	0.353	-0.048	3.019	-0.3478	-3.666	-0.463
(0.001)	(0.000)	(0.002)	0.001)	(0.000)	(0.536)	(0.054)	(0.002)	(0.095)

Table 1 Parameter estimates (full sample 1997-2013, standard errors in parenthesis)

Table 1 presents the estimates of the parameters. Estimated coefficients of the Cobb-Douglas matching function, $\beta_{\log(u)}$ and $\beta_{\log(v)}$, do not confirm the empirical findings that with regional data it may be more likely to find increasing returns in matching (see Ilmakunnas and Pesola [4]). The Czech regional labour market proves the diminishing returns in matching. The elasticity of matches to vacancies is extremely low. That means that the vacancy creation is not a sufficient condition to diminish the unemployment. As the results suggests, new vacancies does not correspond with the structure of the unemployed (e.g. qualification of the unemployed). Elasticity of matching to unemployed, $\beta_{\log(u)}$, shows that approximately half of the new unemployed is able to find a new job immediately. Of course, it is only an approximation.

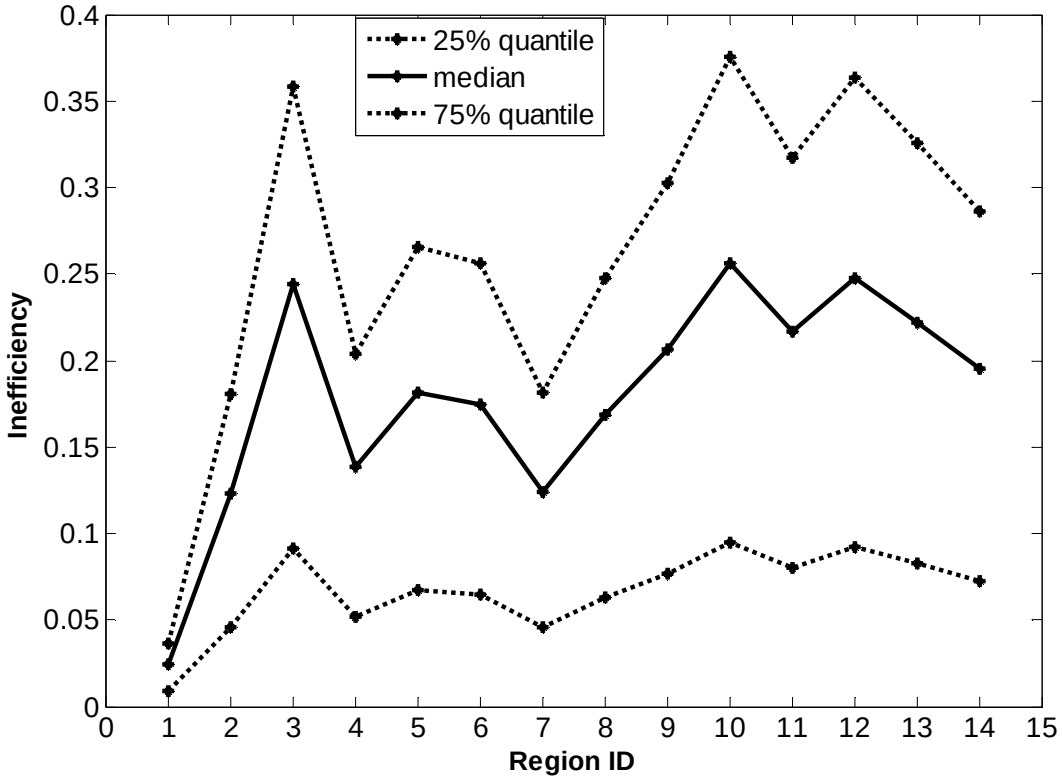


Figure 1 Inefficiency range (full sample 1997-2013)

As for the time trend parameters describing the development of the inefficiency during the period of one year, we can see that the marginal effect of the time on inefficiency is positive but decreasing. Surprisingly, the

GDP growth tends to lower matching efficiency. This may be due to fact, that the positive economic development supports the vacancies creation. But, these vacancies cannot be filled immediately or they are filled by unregistered unemployed. The reason may be in the inappropriate structure of unemployed registered at the employment office. On the other hand, the investment activities lead to the more effective matching. Investment is connected with new vacancies creation and the firms seems to search for the new workers at the employment offices. Moreover, this kind of matching may be supported by the government subsidies. Higher variability, σ_u^2, of the inefficiency term in comparison to the white noise process variability, σ_v^2, contributes to the satisfying identification of the stochastic frontier model (as stated by Wang and Ho [8]).

Figure 1 shows the interquartile range of inefficiency terms distributions for all 14 regions. The identification numbers correspond with those included in the **Table 2**. Minimum inefficiency values are almost zero for all investigated labour markets. It is this clear that all the regions are able to match the unemployed with the vacancies at the full rate. It is caused mostly by the seasonal factors. **Table 2** suggests that there are some regions with exceptionally good performance (Praha or Liberecký kraj) or bad efficiency performance (e.g. Jihočeský kraj or Vysočina). It may be surprising that some discussed regions suffering from high unemployment (e.g. Ústecký kraj) have not the less efficient labour markets. But, it should be stressed that low inefficiency does not automatically means low unemployment. It expresses the potential for new created matches which can be constituted by the interaction between unemployed and available vacancies.

ID	Region	Minimum	25% quantile	50% quantile	75% quantile	Maximum
1	Praha	0.001	0.009	0.025	0.037	0.047
2	Středočeský kraj	0.007	0.046	0.123	0.181	0.230
3	Jihočeský kraj	0.013	0.091	0.245	0.359	0.457
4	Pzeňský kraj	0.007	0.052	0.139	0.204	0.259
5	Karlovarský kraj	0.010	0.068	0.186	0.266	0.339
6	Ústecký kraj	0.010	0.065	0.175	0.256	0.326
7	Liberecký kraj	0.007	0.046	0.124	0.182	0.231
8	Královehradecký kraj	0.009	0.063	0.169	0.247	0.314
9	Pardubický kraj	0.011	0.077	0.207	0.303	0.386
10	Vysočina	0.014	0.095	0.256	0.375	0.478
11	Jihomoravský kraj	0.011	0.081	0.217	0.317	0.404
12	Olomoucký kraj	0.013	0.092	0.248	0.364	0.463
13	Zlínský kraj	0.012	0.083	0.222	0.326	0.415
14	Moravskoslezský kraj	0.010	0.073	0.195	0.286	0.364

Table 2 Regional labour markets inefficiency pattern (full sample 1997-2013)

From this point of view, the results imply that the potential of some problematic labour markets is utilized quite well. There may be an appropriate structure of unemployed and vacancies, unobserved characteristics of the unemployed support their willingness to active job search and finally, the surrounding regions may offer other possibilities for employing unemployed job applicants (this spatial dependency is not implemented in estimated models so far). The unfavourable efficiency outcomes of the regions Vysočina or Olomoucký kraj may be thus explained in a similar way.

Table 2 depicts the distribution of inefficiency terms across the Czech regions during the period from 1998 to 2012. This figure summarise the aggregate regional inefficiency changes in a straightforward way. We can observe relative stable distributions during the whole examined period. The differences across the regions diminished in 2008. These results do not indicate that the estimated labour market inefficiency may rise during the recession and recovery period while it decreases during the economic booms (see Gorter et al. [3]). Although looking at the medians of inefficiency in 2006 and 2009 may suggest some indications of this kind of behaviour.

4 Conclusion

Obtained results shows, that the stochastic frontier model approach is able to capture some interesting patterns of these labour markets controlling individual fixed effects of examined regions and possible time-varying changes in the inefficiency terms. The model estimates using the full sample displays disparities in the labour market inefficiency among the regions although the distribution remain stable through the whole examined period. The

low inefficiency does not necessary mean the low unemployment in the investigated regions. It will be of great importance in further research to focus on the model outcomes with region specific variables. Moreover, the spatial properties of the labour markets dynamics should be investigated, i.e. the efficiency terms should incorporate the influence of neighbouring regions.

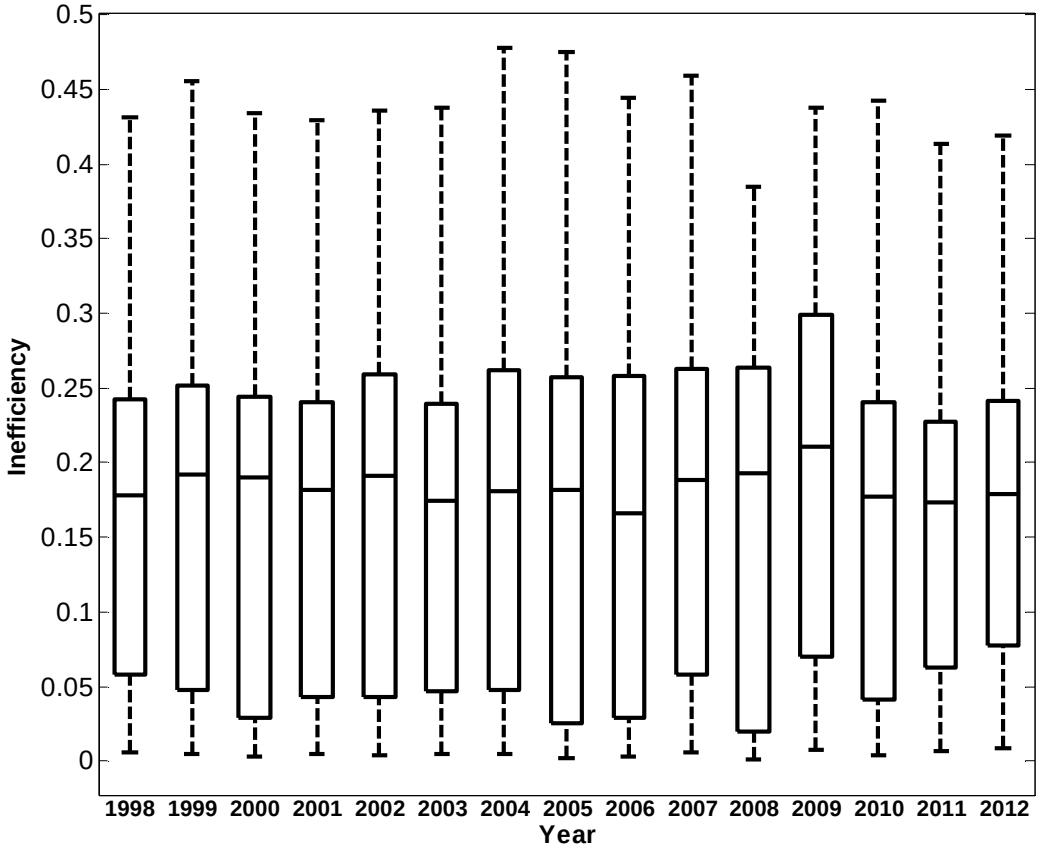


Figure 2 Inefficiency distributions among years (1998-2012)

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0808/2013.

References

[1] Battese, G. E., and Coelli, T. J.: A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics* **20** (1995), 325–332.

[2] Galuščák, K., and Münich, D.: Structural and Cyclical Unemployment: What Can Be derived from the Matching Function. *Czech Journal of Economics and Finance* **57** (2007), 102–125.

[3] Gorter, C., Nijkamp, P., and Pels, E.: Vacancy Dynamics and Labor Market Efficiency in the Dutch Labor Market. *Growth and Change* **28** (1997), 173–200.

[4] Ilmakunnas, P., and Pesola, H.: Regional Labour Market Matching Functions and Efficiency Analysis. *Labour* **17** (2003), 413–437.

[5] Němec, D.: Evaluating labour market flexibility in V4 countries. In: *Proceedings of 31st International Conference Mathematical Methods in Economics* (Vojáčková, H., ed.). College of Polytechnics Jihlava, Jihlava, 2013, 661–666.

[6] Němec, D.: Investigating Differences between the Czech and Slovak Labour Market Using a Small DSGE Model with Search and Matching Frictions. *The Czech Economic Review* **7** (2013), 21–41.

[7] Tvrdoň, M., and Verner, T.: Regional unemployment disparities and their dynamics: Evidence from the Czech Republic. In: *Proceedings of 30th International Conference Mathematical Methods in Economics* (Ramík, J., and Stavárek, D., eds.). Silesian University in Opava, School of Business Administration in Karviná, 2012, 938–943.

[8] Wang, H., and Ho, C.: Estimating fixed-effect panel stochastic frontier models by transformation. Journal of Econometrics **157** (2010), 286–296.

The role of exchange rate dynamics in Bulgaria and Romania in the process of economic transition

Daniel Němec[1], Libor Žídek[2]

Abstract. Our contribution focuses on the role of the exchange rate changes in Bulgaria and Romania during the transition process toward a market economy. We are interested in the degree of exchange rate pass-through to the domestic inflation in these countries. Both of the countries suffered from a high level of inflation and tried to fix their exchange rates in some of the periods. But they were forced to abandon it consequently and it was often followed by sharp depreciation. The goal of our contribution is to evaluate shock absorbing role of the exchange rate changes. We try to verify a traditional hypothesis that exchange rate adjustments are able to accommodate the shocks hitting the economy and to dampen their influence on the other macroeconomic variables. On the other hand, exogenous shocks in the foreign countries may affect exchange rate and lead to additional volatility of the main economic indicators in the domestic economy. This shock generating role of the exchange rate will be evaluated as well. We will use structural vector autoregression models identified by Cholesky decomposition.

Keywords: Bulgaria, Rumania, exchange rate pass-through, structural vector autoregression, Cholesky decomposition.

JEL Classification: C22, F31
AMS Classification: 91B84

1 Introduction

The ex-centrally planned countries decided to follow different exchange rate strategies during their transformation process. Our goal is to detect the impact of the exchange development on inflation in Bulgaria and Romania in the period between the beginning of 1992 and the end of 2008. The first date was selected due to availability of data and the second due to the start of the financial crisis that would affect the overall results. We generally pick up the threads of our previous work (Němec and Žídek [8]) where we analysed the impact of exchange rate regime on inflation among the Central European countries.

We suppose that the different exchange rate strategies had different impacts on inflation development. Foremost we expect that currency board (applied in Bulgaria after 1997) should stabilize it. The same should be true about managed floating that has been used in Romania after 1997. And we expect that the impact of the floating regime will be ambiguous because there are usually periods of both appreciation and depreciation. We split the overall analysed period into sub periods according to the applied exchange rate regime. And we use the country specific Structural Vector Autoregression models (SVAR models) to estimate structural shocks. Using SVAR simulation procedures we are able to decompose the development of inflation into contributions of the specific structural shocks. This method allows us to evaluate the impact of the historical exchange rate shocks on inflation. Our approach differs from many other authors (see e.g. Mirdala [6], [7]) due to the fact that they evaluate the exchange rate pass-through to domestic prices using the forecast error variance decomposition. We are thus able to observe the real historical consequences of the exchange rate changes from the beginning of the 1990s. Although the SVAR approach assumes possible interdependencies among all of the modelled endogenous variables, we will focus on one sided causality from nominal exchange rates changes to price level changes. Our results suggest that the nominal effective exchange rate changes in Bulgaria and Romania determine inflation development in an important way.

2 Exchange rates in the transformation process

In this section we describe different development of exchange rates in Bulgaria and Romania. First of all, we mention theoretical exchange rate systems. The two countries used different exchange rate regimes during the

[1] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, nemecd@econ.muni.cz.
[2]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, l_zidek@centrum.cz.

transformation. In the first period, rather chaotic systems were applied in both of the countries. The central banks did not have the situation under control and attempted to fix/control exchange rates with consequent huge devaluations. These periods were interposed by periods of floating. The exchange rate strategies were determined by high inflation that was characteristic for both of the countries. The second exchange rate regime/period was currency board that has been used in Bulgaria since 1997. It became an anchor of the whole system when the economy stabilized after hyperinflation. The third applied exchange rate regime was managed floating that was used in Romania after 1997. The central bank applied de facto a system with bandwidth.

What was the specific development of the exchange rate regimes? We will consider only the main changes in the development. Bulgaria started the transformation process with an attempt at shock therapy (without restrictive policy). However, contrary to other countries that applied this strategy, the country was not able to fix its exchange rate due to insufficient foreign reserves. But the exchange rate was unified, devalued and the Leva became internally convertible at the beginning of 1991 (Dimitrov [3]). Consequently, the country experienced several episodes of high inflation and deep depreciation. The central bank tried several times to defend the currency by applying its currency reserves but with limited success. For example, in March 1994 the central bank gave up its attempts at stabilising the exchange rate and the Leva depreciated by 50% against the Dollar (Bristow [2]). After relatively calm 1995 the bank again tried to keep the exchange rate at an unrealistic level. And the Leva again collapsed in March 1996. At that time the whole economy was in total disarray because Videnov government tried to reapply some elements of the central planning – for example control of prices. At the same the government had highly ambitious spending plans but was not able to finance them from its incomes and the budget ended with an enormous deficit that was covered by monetization. The obvious consequences were hyperinflation, dollarization and evasion of the Leva. The exchange rate depreciated by 90% against the Dollar in 1996 and this slump continued in the first quarter of 1997 when the Leva dropped from 487 to 1588 per dollar. The government, together with the economy, collapsed in spring of that year. The new government (after the elections) applied a stabilization policy. It incorporated currency board against the German Mark (later on the Euro) in the mid-1997. It was backed 100% by the foreign exchange reserves.

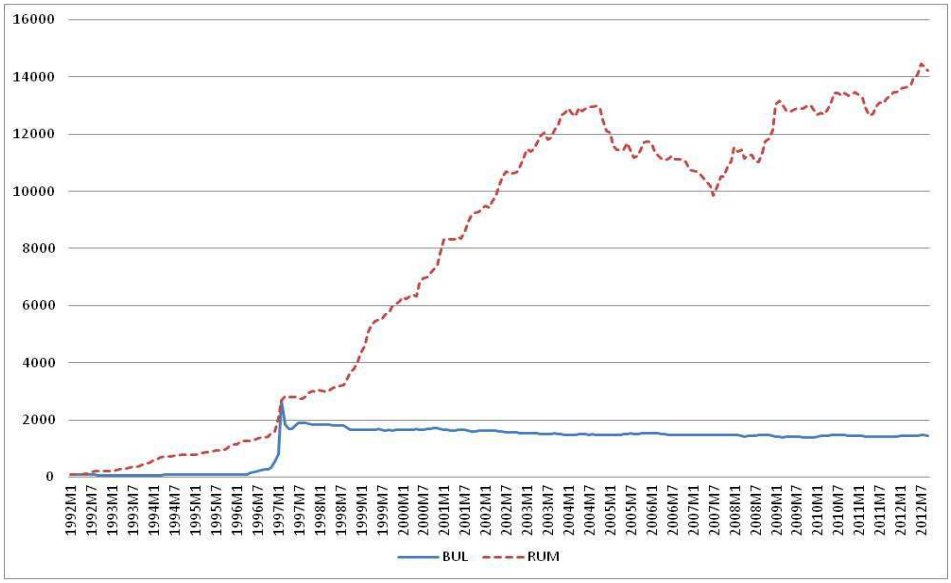


Figure 1 Development of the nominal exchange rates (index, 1992M1 = 100, indirect quote)

The development of Romanian exchange rate policy was no less dramatic and the description is more complicated because opinions among authors vary. The currency was devaluated and unified (the tourist and business exchange rates) in February 1990. There were further devaluations in the following years and the Leu became again operating in the system of multiple exchange rates (Fidrmuc et al. [4]). The situation changed in 1997 when exchange rates were unified and the central bank started to apply managed floating. Fidrmuc et al. [4] write that the system was de facto crawling band with +/- 5% bandwidth. The Leu became convertible for transactions on the current account in 1998 (Popescu [9]). After 2004, the National Bank of Romania relaxed exchange rate policy by decreasing the size and frequency of interventions on the currency market. At the end of the year it further increased the exchange rate flexibility (Penkova-Pearson [10]). These steps were preparatory for a fundamental change in monetary policy because the central bank introduced inflation targeting system in 2005. However it has kept managed float at the same time.

The trend in Romania till 2004 was towards a lasting nominal depreciation. Then the exchange rate relatively stabilized and in some periods appreciated. In Bulgaria the exchange rate was relatively stable (in comparison to

Romania) till the period of hyperinflation that started in 1996. The exchange rate became steady with currency board system after 1997. We can see the development of the nominal exchange rates in Figure 1.[3]

3 Data and methodology

The main goal of our paper is to estimate and quantify the impact of exchange rate changes on domestic inflation. To study the possible factors standing behind the domestic price level changes in the economy we use the methodology proposed by Ito and Sato [5] or Audzei and Brázdík [1]. In order to examine the sources of inflation in the transition period, we will use a structural VAR model of a small economy. This approach allows us to reveal the important interaction between the exchange rate and domestic macroeconomic variables.

Our analysis starts with an estimation of a reduced form VAR model. Based on these estimates we will transform the reduced form residuals into the structural shocks using Cholesky decomposition. That means, we are imposing only the short-term restrictions. As Audzei and Brázdík [1] pointed out, another possibility is to implement sign restrictions where shocks are identified imposing restriction on the signs of the impulse response function to the structural shocks. This approach may solve many puzzles occurring due to alternative estimation procedures. But, regarding our main focus on the transition period of the Romania and Bulgaria, we are aware that traditional sign assumptions about the impulse response functions might be not valid in this case. Cholesky decomposition seems to be the most appropriate and straightforward solution due to the fact that we will use the monthly data and the causal ordering of the variables may thus reflect the instantaneous interdependencies among the variables very well.

Our methodology will be as follows. Identified structural shocks are incorporated into simulation exercises that will decompose the actual trajectories of our observed variables into the particular shock components. The resulting historical shock decomposition reveals the main factors influencing the development of inflation in the transition period. Moreover, we try to quantify the absolute and relative importance of these factors.

Our basic VAR model is set up using the vector x_t of five endogenous variables[4], $x_t = (\Delta OIL_t, gap_t, \Delta m_t, \Delta NEER_t, \Delta p_t)$, where ΔOIL_t denotes changes in the natural log of oil prices, gap_t is the output gap, Δm_t stands for changes in the natural log of money supply, $\Delta NEER_t$ is the growth rate of nominal effective exchange rate and Δp_t is the difference of the natural log of the consumer price index.

The data set used for estimation is from January 1992 to December 2008. The data source is International Financial Statistics (IFS) provided by the International Monetary Fund. The observed variables are as follows:

- IPM: Index of industrial production, 2005=100 (for Romania only);
- TMG: Value of imports, US Dollars (for Bulgaria only);
- CPI: Consumer prices index, all items, 2005=100;
- M1: Monetary aggregate M1, national currency (for Romania only);
- IR: Money market rate, percent per annum (Bulgaria only);
- OIL: Crude oil prices (Brent – Europe), dollars per barrel;
- NEER: Nominal effective exchange rate (index), base year 2005 = 100, direct quote (increase means appreciation).

All variables (except NEER) were seasonally adjusted using the X12-ARIMA procedure. After seasonal adjustment, the variables were transformed using logarithmic transformation. Unit root tests proved the existence of unit roots in all variables. Except IPM and TMP, the variables were expressed in growth rates terms (i.e. logarithmic differences) or changes (interest rate variable). Variables IPM and TMG were filtered using Hodrick-Prescott filter with the smoothing constant $\lambda = 14400$ (this is the standard value for monthly data). This procedure resulted in the corresponding gap_t variable. We prefer to use the gaps as proxy variables for the business cycle. These transformations led to stationary variables. As a monetary policy variable (money supply), the base money (M1) is used for Romania and money market rate (IR) for Bulgaria. As price index variable, we used CPI in our final model. All the variables were selected in accordance with the arguments presented by Ito and Sato [5] and are very similar to those used by Mirdala [7] and Němec and Žídek [8].These variables express economic linkage between inflation and internal and external macroeconomic factors. Oil price inflation represents possible external supply shocks, demand shock effects are included in the output or import gaps, and money supply or changes of the interest rate one the money market allows us to capture the effects of monetary policy on domes-

[3] The data source is International Financial Statistics (IFS) provided by the International Monetary Fund. The observed variables are nominal effective exchange rate (index).

[4] We have checked many other specifications and ordering but the main results presented in this paper remained relatively stable.

tic inflation. We used the import gap for Bulgaria due to the fact, that the index of industrial production was not available till the year 2000. The dynamics of both output gaps is very similar and the import gap may thus serves as proxy variable very well. Similar arguments hold in case of using the interest rates in Bulgaria as a proxy for monetary variable (M1 statistics was available since 1994 only).

Identified structural shocks are based on the Cholesky decomposition of the variance-covariance matrix Ω of the reduced-form innovations. The link between the reduced-form residuals (u_t) and the structural shocks (ε_t) can be written as $u_t = S\varepsilon_t$ where $u_t = \left(u_t^{oil}, u_t^{gap}, u_t^{m}, u_t^{neer}, u_t^{cpi}\right)$, $\varepsilon_t = \left(\varepsilon_t^{oil}, \varepsilon_t^{gap}, \varepsilon_t^{m}, \varepsilon_t^{neer}, \varepsilon_t^{cpi}\right)$ and S is the lower-triangular matrix derived given the covariance matrix Ω. The Cholesky decomposition of Ω implies $\Omega = PP'$ with a lower triangular matrix P. Since $\Omega = E(u_t u_t') = SE(\varepsilon_t \varepsilon_t')S' = SS'$, where structural disturbances ε_t are considered to be orthonormal, the matrix S is equal to P. The structural model defined by the equation (2) is identified due to the fact that the lower-triangular matrix S imposes enough zero restrictions $(k(k-1)/2)$, where k denotes the number of endogenous variables. The lower-triangular matrix S implies that some structural shocks have no instantaneous effects on the chosen endogenous variables. The ordering is thus important and is based on the economic intuition. For example, we assume that changes in oil prices are influenced only by the oil price (supply) shock ε_t^{oil} itself in the first period. As another example, we assume that changes in the output gap may be influenced by the oil price shock immediately, but the changes in monetary policy, exchange rate or inflation have no effect in the same period. These assumptions seem to be intuitive with regards to the monthly data frequency.

4 Estimation results

The lag order of the VAR model was selected using the Akaike information criterion (AIC). In particular, the lag order of 6 periods was selected for Bulgaria and the lag order of 8 periods for Romania. The constant term was included in our model specification. That might be justified by the fact, that there was a significant (i.e. non-zero) growth in the monthly CPI.

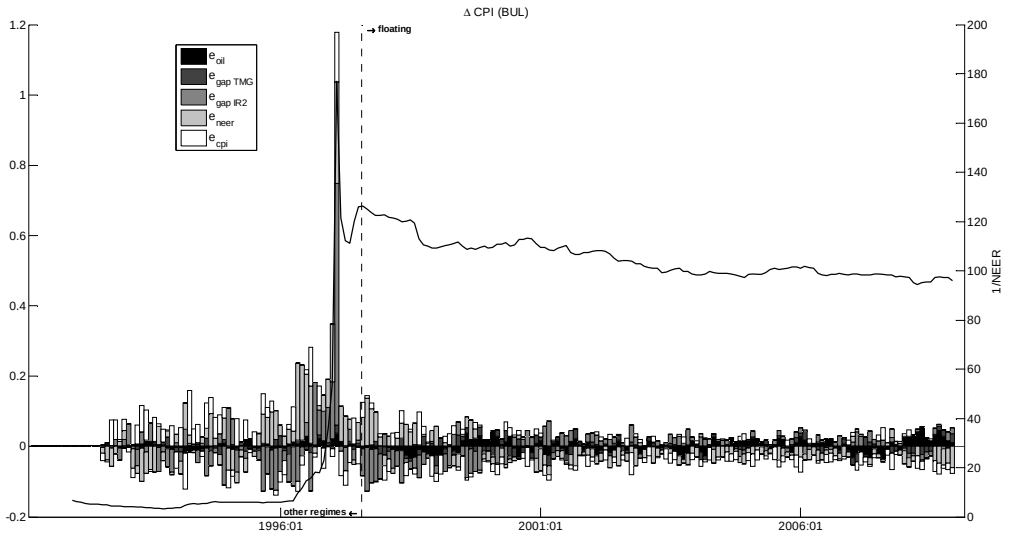


Figure 2 Historical shock decomposition – inflation (Bulgaria)

Figure 2 and Figure 3 show the historical shocks decomposition of monthly CPI inflation and the development of the NEER (indirect quotation where increase means depreciation) for the Romania and Bulgaria respectively. The corresponding structural shocks contributions were computed using the identified SVAR model. The subplots in Figure 2 and Figure 3 contain a mark of the period of change (vertical line) of the exchange rate regimes. These time intervals are used in Table 2 to determine the relative importance of all shocks to the variance in inflation during the examined period. In particular, the start of managed floating exchange rate period was stated in both countries as 1997M6. The previous periods are considered as the regime of various exchange rate combinations including the fixed exchange regime. We can notice (see Figure 2 and Figure 3) that in several periods exchange rate plays an important role in the explanation of inflation. An interesting period is for example the January of 1997 in both countries. The strong devaluations were causing inflation.

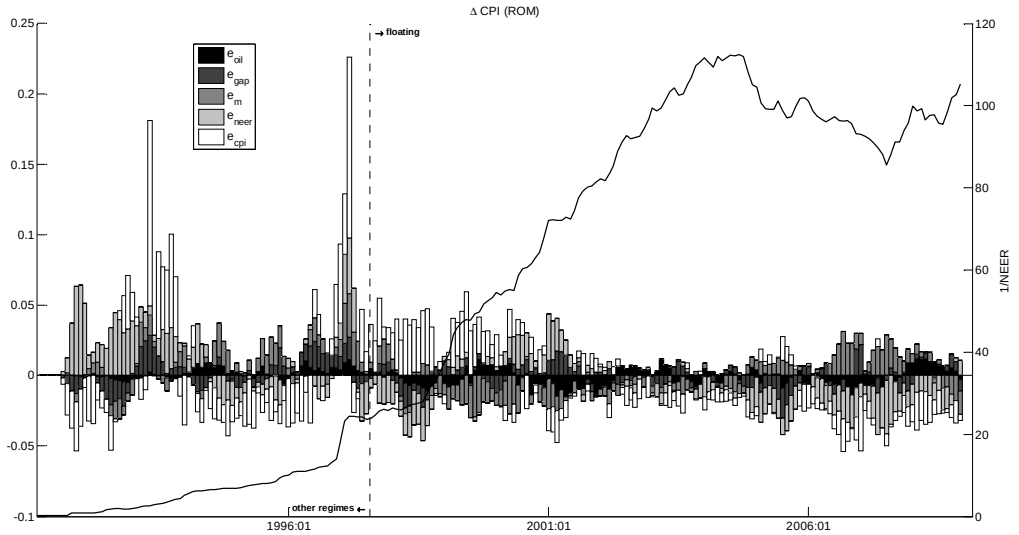


Figure 3 Historical shock decomposition – inflation (Romania)

Forecast error variance decomposition of inflation for both countries is presented in Table 1. The exchange rates are important inflation variables for both countries. Surprisingly, the impact of exchange rate shocks are very similar. The only difference lies in the role of monetary policy. Monetary factors in Bulgaria constitute more than 50% of the realized variability in the inflation. In case of Romania, the half of the inflation variability cannot be explained solely by the structural shocks under consideration.

	ε^{oil}	ε^{gap}	ε^{m}	ε^{neer}	ε^{cpi}
ROM	0.0292	0.0609	0.1012	0.2836	0.5251
BUL	0.0496	0.0159	0.5090	0.2841	0.1414

Table 1 Forecast error variance decomposition of inflation

The evaluation of the relative importance of structural shocks may be found in Table 2. The presented numbers are outputs of the regressions where dependent variable is the standardized inflation and the explaining variables are standardized structural shocks (see Němec and Žídek [8] for further details). There is no violation of classical linear regression assumptions due to the fact, that the explanatory variables are true driving forces for all endogenous variables in our SVAR models. This regression procedure is thus a useful tool for deriving the relative importance of mutually independent structural shock variables. All the computations were carried out for the periods of mixed exchange rate regimes, floating exchange rate and the period of both regimes.

	"1990s"		**"2000s"**		**"1990s" and "2000s"**	
	ROM	**BUL**	**ROM**	**BUL**	**ROM**	**BUL**
	Mixed	**Mixed**	**Float**	**Float**	**All regimes**	
Period	1992-1997	1993-1997	1997 - 2008		1993 - 2008	
ε^{oil}	0.04%	2.73%	4.56%	3.09%	2.01%	3.38%
ε^{gap}	5.95%	1.80%	7.42%	2.29%	3.97%	1.67%
ε^{m}	1.16%	57.26%	5.26%	55.22%	4.49%	47.77%
ε^{neer}	41.05%	27.08%	24.89%	26.09%	45.37%	31.60%
ε^{cpi}	45.29%	10.37%	57.79%	13.33%	35.34%	14.72%

Table 2 Explained historical variability of inflation

For the analysis of the results we divided our period as mentioned above. For easier recognition we roughly divided the period to the 1990s and the 2000s in Table 2. We have computed contribution of the nominal effective exchange rates to the year-on-year inflation in V4 countries. These computations are presented in Table 3. The periods are similar to those from Table 2. The column "Model" contains estimated intercept in the ΔCPI (i.e. inflation) equation of the SVAR model. This estimate corresponds to a part of the inflation which is not explained by the variability of the endogenous variables and may be thus considered as the overall trend in price

level development. We can see the inflationary impacts of the NEER changes in the 1990s and the mostly deflationary tendencies caused by the NEER changes in the 2000s.

	"1990s"		**"2000s"**		**"1990s and 2000s"**		**Model**
	ε^{neer}	ΔCPI	ε^{neer}	ΔCPI	ε^{neer}	ΔCPI	ΔCPI
ROM	4.15%	98.48%	-6.58%	21.40%	-3.68%	39.40%	31.16%
BUL	73.64%	127.63%	-14.62%	6.71%	5.65%	33.94%	26.77%

Table 3 Contribution of NEER changes to the average year to year CPI inflation

The goal of our contribution is to evaluate shock absorbing role of the exchange rate changes. Table 3 shows that in case of Bulgaria, exchange rate shocks contributed very strongly to the inflation pressures in the first half of 1990s. The exchange rate changes contributed to the average year on year CPI inflation of 127% by increasing the price level by 74% in this period. On the other hand, the high average year on year CPI inflation of 99% was not determined by the exchange rate movements. The average contribution of 4% was thus negligible compared to the average 100% price level changes. After adoption of managed floating, both countries experienced negative impact of exchange rate changes on domestic inflation. That was crucial for the mild price level development in Bulgaria.

5 Conclusion

In our contribution, we have analysed the relationship between exchange rates and inflation in Bulgaria and Romania. These countries applied different exchange rate strategies during the transformation process. We have found out that change in exchange rates played an important role in the first half of 1990s. We can conclude that the exchange rate adjustments were not able to accommodate the shocks hitting both economies and to dampen their influence on the inflation in the first half of the 1990s. Exogenous shocks affected exchange rate and led to additional volatility of the main economic indicators in the domestic economy. After 1997, both countries changed their exchange rate regimes (to the managed floating and currency board). As a result, the exchange rate adjustments accommodated the shocks affected the CPI inflation in this period.

Acknowledgements

This work is supported by funding of specific research at Masaryk University (project MUNI/A/0811/2013). We also thank to the anonymous reviewer for detailed corrections, comments and suggestions.

References

[1] Audzei, V. and Brázdík, F.: Monetary Policy and Exchange Rate Dynamics: The Exchange Rate as Shock Absorber. Czech National Bank Working Paper Series 9, 2012.
[2] Bristow, J.: Bulgaria in The Central & Eastern Europe Handbook, Prospects onto the 21st Century, edited by P. Heenan, M. Lamontage, Fitzroy Dearborn Publishers, London, Chicago, 1999.
[3] Dimitrov, V.: Bulgaria the uneven transition, Routledge, London, 2001.
[4] Fidrmuc, J., Silgoner, M. A., and Cuaresma, J. C.: Exchange Rate Developments and Fundamentals in Four EU Accession and Candidate Countries: Bulgaria, Croatia, Romania and Turkey. *Focus on European Economic Integration*, issue 2, Oesterreichische Nationalbank (Austrian Central Bank).
[5] Ito, T., and Sato, K.: Exchange Rate Changes and Inflation in Post-Crisis Asian Economies: Vector Autoregression Analysis of the Exchange Rate Pass-Through. *Journal of Money, Credit and Banking* **40** (2008), 1407–1438.
[6] Mirdala, R.: *Menové kurzy v krajinách strednej Europy*. Technická univerzita v Košiciach, Košice, 2011.
[7] Mirdala, R.: Exchange rate pass-through to domestic prices in the Central European countries. *Journal of Applied Economic Sciences* **4** (2009), 408–424.
[8] Němec, D., and Žídek, L.: The impact of exchange rate changes on inflation in the V4 countries in the process of economic transition. In: *Proceedings of the 31st International Conference Mathematical Methods in Economics 2013* (Vojáčková, H. ed.). College of Polytechnics Jihlava, Jihlava, 2013, 1087–1092.
[9] Popescu, D.: Internal Convertibility and Liberalization in the Transition Period: The Romanian Case in Impediments to the Transition in Eastern Europe edited by M. Uvalic, E. Espa, J. Lorentzen, European University Institute, Florence, 1993.
[10] Penkova-Pearson, E. (2011). *Trade, convergence and exchange rate regime: Evidence from Bulgaria and Romania.* Sofia, 2011.

SDP application on portfolio optimization problem with non-convex quadratic constraints

Kirill Odintsov[1]

Abstract. This paper is focused on the problem of portfolio optimization. This problem is formulated as a mean-risk model with a special choice of shortfall constraints that leads to quadratic programming. We show the methodology of finding an optimal solution to such problem even if the said quadratic programming is non-convex. The non-convexity comes from our demand that the model should favour the positive deviations from expected returns of the portfolio. First the former problem is converted to semi-definite programming (SDP) problem using semi-definite relaxation technique. In the literature it is shown that this will provide a good proximate for the former problem. The SDP is solved using interior point methods with self-concordant barrier. These methods have in worst case scenario polynomial time complexity. The last issue that we face is the issue of feasibility of the solution of the relaxed problem. We propose using eigenvalue method and randomization method to deal with this issue. The whole process is demonstrated on a real data. It is implemented in R using *Rcsdp* package.

Keywords: Semi-definite relaxation of quadratic programming, portfolio optimization, interior point method

JEL classification: C44
AMS classification: 90C20

1 Introduction

Mean-risk models are drawing a lot of attention since they were introduced in [5] for portfolio optimization. Since introduction there have been many improvements in this field. Some of the the research is dedicated to finding the appropriate risk measures like for example value-at-risk, conditional value-at-risk and other alternatives. The lately discussed topic is to use models with second order stochastic dominance as an alternative to mean-risk models. This idea originated in [2] and continued in [3]. The stochastic dominance approach has two advantages over the mean-risk models which are the fact that there is no need for a subjective choice of the risk measure and that the optimal solution found this way can't be dominated by any other solution. The disadvantage is higher computational and theoretical complexity.

This article is limited on such mean-risk models that lead to quadratic optimization. The problem of portfolio optimization with the linear costs and special choice of shortfall constraints leads to convex quadratic optimization which is effectively solved by standard methods of convex optimization. In this article we show the solution that applies even for those mean-risk models that lead to non-convex quadratic programming. There are several articles that offers solution to such problem as for example [6]. But we haven't found one that uses SDP relaxation method. The advantage of using semi-definite programming is that it offers a possibility to find a good approximate solution to many hard problems in polynomial time. We applied general methodology described in [7] to the specific case of portfolio optimization. In our case more general choice of shortfall constraints is the source of non-convexity. First we convert the problem to semi-definite programming form using semi-definite relaxation technique as in [7]. In [9] it is shown that this will provide a good proximate for the former problem. The SDP problem has in worst case scenario polynomial time complexity as shown in [9]. The feasibility issue of the solution of relaxed problem is solved using eigenvalue method and randomization method.

The main purpose of this paper is to demonstrate how the SDP relaxation method can be practically implemented on the special case of portfolio optimization problem. This is why we show the whole process

[1]VŠE, FIS, Department of Econometrics, W. Churchill sq.4, Prague, CZ, kodintsov@hotmail.com

on a sample portfolio in part 4 of the text.

2 Portfolio optimization problem formulation

Lets assume that our goal is to construct an optimal portfolio. We can choose from n asset. We denote $x = (x_1, \ldots, x_n)^T$ the weights of the asset in the portfolio. $R = (R_1, \ldots, R_n)^T$ are random variables representing the returns of the assets in the future. The basic form of mean-risk model introduced in [5] would be

$$\begin{aligned} &\min_{x \in \mathbb{R}^n_+} -E[R]^T x \\ &\text{under constraints} \\ &x^T \operatorname{var}[R] x \leq v_1 \\ &e^T x = 1. \end{aligned} \tag{1}$$

where e is appropriate vector of ones, v_1 is the maximal variance that is tolerated by the investor, $E[.]$ denotes expected value and var[.] denotes variance. We of course have to assume that random vector R has finite second moments.

We assume that we have T historic observations of the returns of considered assets. If we replace the vector of expected values and covariance matrix in (1) by the sample mean vector $\overline{R}$ and sample covariance matrix $\overline{V}$ we will be faced with a convex quadratic programming problem. One of the disadvantages of this model is that it treats risk represented by $x^T \overline{V} x$ only as something negative, neglecting that $x^T \overline{V} x$ is also a chance of better than expected returns (this model penalizes under and over performance equally). To capture investors inclination to better than expected returns some authors suggest adding third moment (in form of skewness) to the model. In this paper we would like to focus on quadratic programming so instead of adding the skewness we add the lower bound to the variance constraint

$$v_2 \leq x^T \operatorname{var}[R] x \tag{2}$$

to the model (1). This modified model can be interpreted that the investor still wants the risk to be under certain level but he doesn't want portfolio with too low variance for the possibility of higher returns than expected value. Note that the condition (2) is no longer leading to convex problem.

Now we assume that the R has n-dimensional normal distribution with mean $\overline{R}$ and variance matrix $\overline{V}$. Then we can impose that the portfolio returns are not lower than some fixed value s with some probability $\eta \in [0, 1]$ by constraint

$$P(R^T x \geq s) \geq \eta.^1 \tag{3}$$

Using the information about distribution of R the constraint (3) can be expressed as

$$\overline{R}^T x - \phi^{-1}(\eta) \sqrt{x^T \overline{V} x} \geq s, \tag{4}$$

where ϕ^{-1} denotes quantile function of normalized normal distribution. We will add two versions of constraint (4) to the model. One will be with low s and high η (higher than 0.5) ensuring that the probability that the returns are not too low is high. The other one will have high s and low η (lower than 0.5) giving us some opportunity to have better returns than expected. According to [6] the high choice of η defines a convex constraint but the low choice does not.

Combining constraints (4) and (2) with (1) and adding a non-negative variable $\sigma := \sqrt{x^T \overline{V} x}$ and associated constraint $\sigma^2 = x^T \overline{V} x$ we get the final model

$$\begin{aligned} &\min_{x \in \mathbb{R}^n_+, \sigma \in \mathbb{R}_+} -\overline{R}^T x \\ &\text{under constraints} \\ &\sigma^2 = x^T \overline{V} x, \quad \sigma^2 \leq v_1, \quad \sigma^2 \geq v_2 \\ &\overline{R}^T x - \phi^{-1}(\eta_l)\sigma \geq s_l, \quad \overline{R}^T x - \phi^{-1}(\eta_h)\sigma \geq s_h \\ &e^T x = 1. \end{aligned} \tag{5}$$

(5) is a quadratic non-convex model.

[1] Note that $R^T x$ is a random variable since R is random vector

3 Solution of the problem using semi-definite relaxation

In this section we will apply the theory from [7] to our specific case. First we will convert (5) to homogeneous non-convex quadratic form. Then we will apply the rank relaxation to get semi-definite problem. This problem will be solved using interior-point methods. Afterwards a solution of the former problem (5) will be derived.

3.1 Conversion to homogeneous quadratic non-convex problem

The model (5) is not homogeneous quadratic problem. Let us convert it to the homogeneous quadratic problem by adding a variable t and one extra constraint $t^2 = 1$. This method is suggested in [7]. The homogeneous version of model (5) looks as follows.

$$\begin{aligned}
&\min_{z\in\mathbb{R}^{n+2}} z^T C z \\
&\text{under constraints} \\
&z^T A_1 z = 0, \quad z^T A_2 z \le v_1, \quad z^T A_2 z \ge v_2 \\
&z^T A_3 z \ge s_l, \quad z^T A_4 z \ge s_h \\
&z^T A_5 z = 1 \\
&z^T B z = 1 \\
&z^T M_r z \ge 0 \quad \text{for } r = 1, \dots, n+1
\end{aligned} \tag{6}$$

where

$$z = \begin{bmatrix} x \\ \sigma \\ t \end{bmatrix}, C = \begin{bmatrix} 0 & \cdots & 0 & 0 & -\overline{R}_1/2 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & -\overline{R}_n/2 \\ 0 & \cdots & 0 & 0 & 0 \\ -\overline{R}_1/2 & \cdots & -\overline{R}_n/2 & 0 & 0 \end{bmatrix}, A_1 = \begin{bmatrix} \overline{V} & \mathbf{0} & 0 \\ \mathbf{0}^T & -1 & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

$$A_2 : (A_2)_{ij} \begin{cases} 1 & \text{for } i = j = n+1 \\ 0 & \text{else} \end{cases}$$

$$A_3 = \begin{bmatrix} 0 & \cdots & 0 & 0 & \overline{R}_1/2 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \overline{R}_n/2 \\ 0 & \cdots & 0 & 0 & -\phi^{-1}(\eta_l) \\ \overline{R}_1/2 & \cdots & \overline{R}_n/2 & -\phi^{-1}(\eta_l) & 0 \end{bmatrix}, A_4 = \begin{bmatrix} 0 & \cdots & 0 & 0 & \overline{R}_1/2 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \overline{R}_n/2 \\ 0 & \cdots & 0 & 0 & -\phi^{-1}(\eta_h) \\ \overline{R}_1/2 & \cdots & \overline{R}_n/2 & -\phi^{-1}(\eta_h) & 0 \end{bmatrix}$$

$$A_5 = \begin{bmatrix} 0 & \cdots & 0 & 0 & 1/2 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 1/2 \\ 0 & \cdots & 0 & 0 & 0 \\ 1/2 & \cdots & 1/2 & 0 & 0 \end{bmatrix}, B : (B)_{ij} = \begin{cases} 1 & \text{for } i = j = n+2 \\ 0 & \text{else} \end{cases}$$

$$M_r : (M_r)_i j = \begin{cases} 1/2 & \text{for } i = n+2, j = r \\ 1/2 & \text{for } i = r, j = n+2 \\ 0 & \text{else} \end{cases}, \quad \text{for } r = 1, \dots, n+1$$

All the matrices are symmetric and are from $\mathbb{R}^{n+2\times n+2}$. $\mathbf{0}$ denotes an appropriate vector of zeroes. Matrices $C, A_1, \dots A_5$ correspond to the constraints that were explicitly written in model (5). Matrix

B corresponds to the constraint $t^2 = 1$ and $M_1, \dots M_n + 1$ represents non-negativity constraints of parameters $x_1, \dots, x_n + 1, \sigma$. Note that model (6) and (5) are equivalent in a way that if we find solution $z^* = (x^{T*}, \sigma^*, t^*)$ to (6) then for $t^* = 1$ (x^{T*}, σ^*) is optimal solution of (5) and if $t^* = -1$ the $(-x^{T*}, -\sigma^*)$ is optimal solution of (5).

3.2 Conversion to SDP

Using a simple property of trace of matrices that applies to any two general matrices F, G with appropriate size

$$\text{trace}(FG) = \sum_i (FG)_{ii} = \sum_i \sum_j F_{ij} G_{ji} = \sum_j \sum_i G_{ji} F_{ij} = \sum_i (FG)_{jj} = \text{trace}(GF)$$

and the fact that trace of a number is the number itself we can rewrite the objective function and all the constraints of (6) as

$$z^T \text{"Matrix"} z = \text{trace}(z^T \text{"Matrix"} z) = \text{trace}(\text{"Matrix"} z z^T).$$

Now as in [7] we substitute $Z = zz^T$. The new constraint $Z = zz^T$ is equivalent to two constraints: Z is positive semi-definite and rank(Z)=1. This leads to new model

$$\begin{aligned}
&\min_{Z \in \mathbb{S}^{n+2}} \text{trace}(CZ) \\
&\text{under constraints} \\
&\text{trace}(A_1 Z) = 0, \quad \text{trace}(A_2 Z) \le v_1, \quad \text{trace}(A_2 Z) \ge v_2 \\
&\text{trace}(A_3 Z) \le s_l, \quad \text{trace}(A_4 Z) \le s_h \\
&\text{trace}(A_5 Z) = 1 \\
&\text{trace}(BZ) = 1 \\
&\text{trace}(M_r Z) \ge 0 \quad \text{for } r = 1, \dots, n+1 \\
&\text{rank}(Z) = 1, \quad Z \succeq 0,
\end{aligned} \tag{7}$$

(7) almost looks like a semi-definite problem. There is only one extra constraint - the rank constraint. So we relax the problem by dropping the rank constraint to get the model

$$\begin{aligned}
&\min_{Z \in \mathbb{S}^{n+2}} \text{trace}(CZ) \\
&\text{under constraints} \\
&\text{trace}(A_1 Z) = 0, \quad \text{trace}(A_2 Z) \le v_1, \quad \text{trace}(A_2 Z) \ge v_2 \\
&\text{trace}(A_3 Z) \le s_l, \quad \text{trace}(A_4 Z) \le s_h \\
&\text{trace}(A_5 Z) = 1 \\
&\text{trace}(BZ) = 1 \\
&\text{trace}(M_r Z) \ge 0 \quad \text{for } r = 1, \dots, n+1 \\
&\quad Z \succeq 0,
\end{aligned} \tag{8}$$

where $\mathbb{S}^{n+2}$ denotes the space of symmetric matrices of size $n + 2 \times n + 2$ and $Z \succeq 0$ denotes that Z is positive semi-definite matrix. The used relaxation is well described in literature. In [9] chapter 13 is dedicated to semi-definite relaxation of non-convex quadratic program. There are other relaxation techniques how to transform general non-convex quadratic programming to SDP. These methods are also described in [9] and in [1]. We have chosen the rank relaxation method since it was suggested in [7]. As said in [7] and shown in [9] for some special cases the optimal value of objective functions of (7) and (8) are close.

The problem (8) can be solved by interior-point methods in polynomial time as shown in [8] or [9]. Let us assume that we have found an optimal solution Z^* of (8). If $\text{rank}(Z^*) = 1$ then the optimal solution to (7) and thus to (5) is

$$x_i^* = \sqrt{(Z^*)_{ii}} \quad \text{for } i = 1, \dots, n.$$

Note that in this case it is not a proximate.

If $\text{rank}(Z^*) \neq 1$ then we have a problem that Z^* is not feasible solution of (7), but as mentioned before it is proven that the optimal value of objective functions of both problem are close. So we can search for feasible solution from all the solutions that have approximately same value of objective function as (8). Other option to deal with this problem is to take a rank one approximation

$$\widehat{Z} = \lambda_1 q_1 q_1^T$$

of Z^*, where q_1 is eigenvector and λ_1 eigenvalue. If this is a feasible solution of (7) then the solution of (6) is

$$\widehat{z} = q_1 \sqrt{\lambda_1}. \tag{9}$$

In case $\widehat{Z}$ is not a feasible solution of (7) then we have to search a feasible solution of (6) as close to Z^* as possible. Other option of getting the solution is to use so called randomization method. This method is described in [7]. In this paper we will only describe the method itself without proof or motivation to why and how it works for more details please look at [7]. Generally the method works as that we generate K random vectors ξ of size $n+2$ with distribution $N(0, Z^*)$. Then we create K potential solutions $z(\xi)$ as functions of ξ so these solutions are feasible solutions of the former quadratic problem. Afterwards we compute the objective function with all $z(\xi)$. The proximate solution to our problem will be the one that has the lowest value of objective function. In our specific case we have too many different constraints so the creation of an universal transformation i very hard. Instead for simplification we will just use

$$z(\xi)_i = \begin{cases} \frac{|\xi_i|}{\sum_{j=1}^{n} |\xi_j|} & \text{for } i = 1, \ldots, n \\ x(\xi)^T \overline{V} x(\xi) & \text{for } i = n+1 \\ \text{sign}(\xi_i) & \text{for } i = n+2 \end{cases}. \tag{10}$$

We do realize that the solutions that are obtained this way will still not be feasible in many case, but as mentioned before finding the any transformation that would make feasible $z(\xi)$ is very hard. In this case we have to first test the feasibility of $z(\xi)$ and then proceed to calculate the objective values for feasible $z(\xi)$. Let us note that for this method to give reasonable results we have to select quite high K. So by not locating the hard transformation we paid by increased computational difficulty of the problem.

4 Practical application

The methodology was applied on portfolio with stock of 6 ($n = 6$) companies: China Petroleum & Chemical Corp., Delta Air Lines Inc., The Walt Disney Company, International Business Machines Corporation, Microsoft Corporation and Unilever.

The data are monthly data from 1.1.2008 to 1.1.2013. The companies were chosen so that our portfolio is diversified. The methodology was applied only on 6 companies for easier application. Adding extra companies would not be difficult from computational point but the choice of companies would be and the focus of this paper is mainly on optimization method for portfolio optimization not the preliminary selection of potential assets for portfolio.

After looking at the data and its mean and covariance matrix we selected the parameters of the model as follows

$$v_1 = 0.02, v_2 = 0.001, \eta_l = 0.01, s_l = 0.01, \eta_h = 0.95, \eta_h = -0.5$$

so that the model has non-empty feasible set.

The computation was made in *R* package *Rcsdp*. A slight modification had to be made to the matrices of (8) since the package only supports constraints in form of equivalence. We had to add p slack dimensions to our matrices where p was number of equations with inequalities. Lets again denote the solution of model (8) as Z^* the value of objective function of SDP model (8) for this solution is **0.01303054**.

The rank of Z^* is 8 so it is not feasible solution of (7). We compute the eigenvectors and eigenvalues of Z^* and use (9) to compute an approximate solution of (7). This approximation proved to be infeasible. So at last we search for the solution former problem (7) using randomization method. We simulate 1000 000 times ξ from 8 dimensional normal distribution $N(0, Z^*)$. Then we transform every simulation using (10) getting 1000 000 $z(\xi)$. Afterwards we check which $z(\xi)$ are feasible solutions of former problem (7).

It turns out 585 893 are indeed feasible. From all these feasible solution we take the one that minimizes the objective function of (7) as a solution. Leaving out two last variables we get our close to optimal solution of former problem (5) which is

$$x^* = (0.006421189, 0.008303971, 0.302064784, 0.678157826, 0.001000306, 0.004051923)$$

with value of objective function **0.01288698**.

Note that as a theory suggested this objective function value is very close to the optimal objective function value of (8) (which was 0.01303054).

5 Conclusion

We applied the methodology successfully on the real data. But since a lot of simulation and testing feasibility were involve we surely did not solve the problem in polynomial time. In order for this method to be effective further work has to be done on improving the transformation (10). Other interesting improvement would be designing custom made interior-point method for SDP in form of (8), that would be more effective and would prefer the solution with lower rank of matrices. Other extension of this paper would be to compute the proximity of the (7) and (8) for this specific problem. Another interesting financial application of SDP to further look into are asset and liability problems. As described in [4] when we use the scenario tree generation to solve these problems a non-convex quadratic programming problem arises.

Acknowledgements

This paper supported by the grant No. 58/2014, IG403024 of the Internal Grant Agency. In this place I would like to thank Vladimír Holý who actively advised me and helped me with the software implementation of SDP.

References

[1] Bomze, I. M., Klerk, E.: Solving Standard Quadratic Optimization Problems Via Linear, Semidefinite and Copositive Programming, *J. Global Optim* **24** (2001), 163–185.

[2] Dentcheva, D., Ruszczynski, A.: Optimization with stochastic dominance constraints, *SIAM Journal on Optimization* **14** (2003), 548–566.

[3] Dentcheva, D., Ruszczynski, A.: Portfolio optimization with stochastic dominance canstraints, *Journal of Banking and Finance* **30** (2006), 433–451.

[4] Gondzio, J., Grothey, A.: Parallel Interior-Point Solver for Structured Quadratic Programs: Application to Financial Planning Problems *Annals of Operations Research* **152** (2007), 319–339.

[5] Markowitz, H.: Portfolio selection. *The Journal of Finance* **7** (1952), 77–91.

[6] Lobo, M. S., Fazel, M., Boud, S.: Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research* **152** (2007), 341–365.

[7] Luo, Z.-Q., Ma, W.-K., So, A.M.-C., Ye, Y., Zhang, S.: Semidefinite Relaxation of Quadratic Optimization Problems. *IEEE Signal Processing Magazine* **27** (2010), 20–34.

[8] Porkolab, L., Khachiyan, L.: On the Complexity of Semidefinite Programs. *Journal of Global Optimization* **10** (1997), 351–365.

[9] Wolkowicz, H., Saigal, R., Vandenberghe, L.,: *Handbook of Semidefinite Programming - Theory, Algorithms, and Applications*. Springer, 2000.

Applying System Dynamics and Agent-based Modelling in Archaeological Simulation

Kamila Olševičová[1] and Alžběta Danielisová[2]

Abstract. We apply the social simulation to the cultural milieu of the late Iron Age in Central Europe and its economic strategies, especially on the subsistence and the carrying capacity of the settlement agglomeration in relation to the daily economic needs of its inhabitants. By means of the computational models of the economics of the particular central settlement (Staré Hradisko oppidum) and its hinterland we aim to verify hypotheses about the economic sustainability in case self-supplying is applied. To achieve this, the agent-based modelling, cellular automata approach and system dynamics modelling are used. In the paper we present four alternative scenarios of population dynamics (baseline and three versions of depopulation) and four alternative scenarios of food production and agricultural practices (baseline and three types of external events). The set of models for all scenarios was implemented in NetLogo and Stella. Outputs of the models are subject of interests of archaeologists and according to their feedback the models will be refined and extended.

Keywords: agent-based model, cellular automata, land use model, NetLogo, social simulation, Stella, sustainability, system dynamics.

JEL Classification: C63
AMS Classification: 68U20

1 Introduction

Complex agent-based models and simulations help us to better understand the functioning and the sustainability of socio-economic systems. The explanatory computational models can build on different approaches and methods including system dynamics, agent-based models, cellular automata or networks-based models. The key issue is the availability of data and sources of knowledge about the society in question and its economic system, i.e:

- completeness of archaeological record,
- reliability of written resources,
- broader acceptance of different archaeological and historical hypotheses.

The more resources are available, the more comprehensive models can be developed. Highly organized societies with detailed written official sources (registers of births and deaths, accounting books, chronicles or memoirs) enable creation of simulations of day-by-day functioning of thousands of inhabitants living in large areas. See e.g. the ENKIMDU simulation framework prototype of Early Bronze Age Mesopotamia settlement systems in irrigated regions (see [1],[11]) or the proof-of-concept MayaSim model of ancient Maya social-ecological system (see [5]).

Our research is focused on the explanation of the collapse of Celtic society in Central Europe in late Iron Age using computational simulation of daily economic needs and economic activities of population of the settlement agglomerations ("oppida"). The archaeological record shows that dynamics of settlement occupation includes fast growth followed by rapid decline: the population density peaked within 70 years, and then, within two generations it decreased extensively. This assumption is based on the archaeological evidence given by chronologically significant artefacts such as coins, ceramics, brooches and animal bones in settlement layers from various oppida locations (Staré Hradisko, Stradonice, Závist, Hrazany and others). According to [3], the crisis factors could be political, organizational, economic (long distance contacts) or ecological (i.e. subsistential). The method of social simulation is expected to provide better insight in the situation in relation to the current data and theories. Previous research results were summarised in [2].

[1]University of Hradec Králové, Faculty of Informatics and Management, Dept. of Information Technologies, Rokitanského 62, Hradec Králové 500 03, Czech Republic. kamila.olsevicova@uhk.cz.
[2]Institute of Archaeology of Academy of Sciences of the Czech Republic, Letenská 4, 11801, Prague 1, Czech Republic. danielisova@arup.cas.cz.

The paper presents the population dynamics models and scenarios (section 2) and the subsequent agricultural models (section 3). The following research intentions are mentioned in the Conclusion.

2 Population dynamics scenarios and models

The population dynamics modelling has got two inputs: initial population which is between 600-800 individuals in 2 sex and 7 age categories (suckling, toddlers, children, older children, young adults, adults and elderly) and model life tables for ancient Roman female population which were identified by experts as the most relevant resource of information about the birth-rate and the death-rate of the ancient population.

The model is designed to provide time series of the synthetic population data (i.e. numbers of individuals in individual age/sex groups, number of "strong" (i.e. adult males) and "weak" (i.e. females, children and elderly) workforce and the consumption of the population in calories) for the period of 120 years. The constraints to be satisfied are especially the constant growth rate, approx. constant ratio of male/female and approx. constant ratio of age categories.

The first version of the model was implemented in NetLogo using the agent-based approach. Each individual is represented by agent. The simple life cycle of agents is defined using the approximated Roman life tables. Agents representing females between 15 and 45 years of age apply the birth-rate procedure. Details are provided in [7] and [9].

The second version of the model was implemented in NetLogo and Stella using the system dynamics approach. The age structure of the population is represented by the system of stocks and flows. Both implementations were compared in [10]. For Stella simulation see Fig. 1: the quickly growing young populations are typically characterized by slow decline of average age.

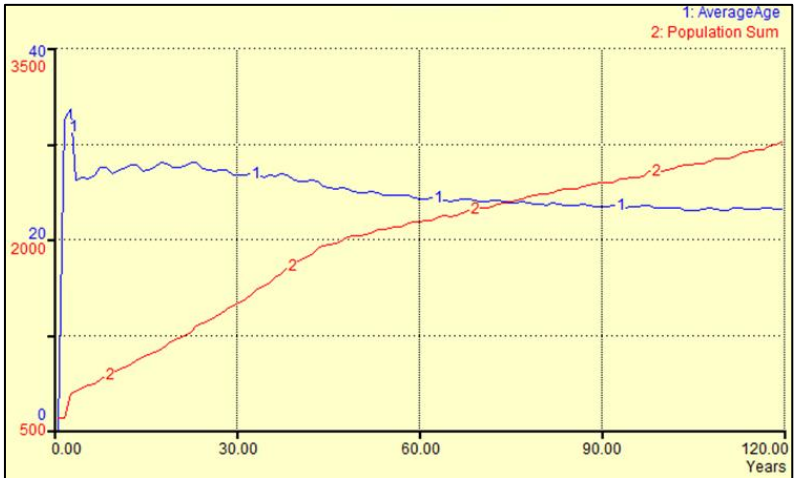


Figure 1 Synthetic population in Stella

The third version of the model generates synthetic populations within four predicted scenarios:

- Normal growth (*baseline*) – the population grows from 600-800 up to 2500-3000 individuals, outputs are identical with previous versions of the model.
- Sudden proportional decline (*sudden depopulation*) – massive one-time depopulation of 30-60 percent of inhabitants. It is naturally accompanied with the decline of workforce, livestock and food storage. It corresponds to the hypothesis of the moving of the part of the population to the more suitable territory.
- Sudden non-proportional decline (*epidemic*) – massive one-time depopulation of 30-60 percent of inhabitants, more devastating in certain age groups (suckling, toddlers, children, elderly). It causes significant workforce decline in following decades.
- Continuous proportional decline (*gradual depopulation*) – less intensive but continuous depopulation. It corresponds to the hypothesis of continuous emigration.

The increasing and/or decreasing population trend should be reflected in models of food and fodder production (i.e. the spatial change in the field, pasture, and forest area) as well as in the numbers of livestock. The population dynamics model was implemented in NetLogo and re-implemented in system dynamics in order to test the model´s logic. See Fig. 2 for the simulation outputs in NetLogo: time series of the population in total and in age groups, the stable gender structure and the changing workforces.

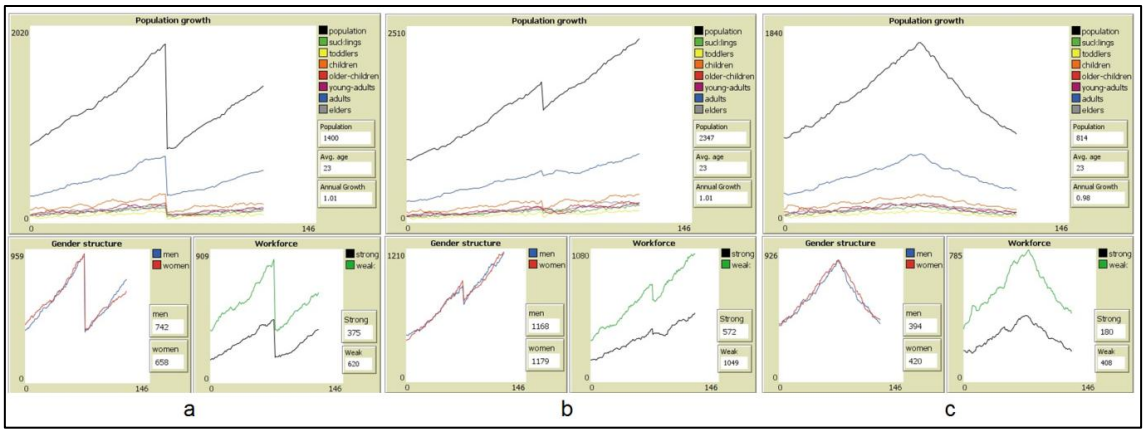


Figure 2 Depopulation scenarios in NetLogo: (a) sudden depopulation, (b) epidemic, (c) gradual depopulation. Graphs show numbers of individuals in seven age groups (top, black = total number of population), gender structure of the population (bottom left, no changes) and numbers of strong-force and weak-force (bottom right).

3 Agricultural practices scenarios and models

The objective of the agricultural practices model is to explore the economic functioning of the settlement, especially covering its subsistence requirements and carrying capacity of the environment. The essential hypothesis supposes the continuous growth the population reflected in proportional increase of the land use. The bigger population has got higher consumption requirements. The more workforce is available, the larger part of the original woodland cover around the settlement can be cut or burnt, changed into the arable land or pastures and used as means for the food production. The growth of the population is naturally limited by sources (human workforce, animal labour availability – such as the ploughing oxen) and the land resources within the oppidum´s hinterland as one of the crucial factors for the prediction of both fields and pastures was the accessibility from the settlement (1 hour of walking distance from the settlement is said to be realistic). Availability of workforce also constraints the land use. If the limits were achieved, the population has either to adapt in order to sustain or decrease. The agricultural strategy can be changed (e.g. intensive growing is replaced with extensive growing having lower workforce requirements), diet (the ratio of cereals, pulses, meat and milk) can be modified, or part of the inhabitants can leave the settlement.

The inputs of the agricultural practices model are more complex than those of the population dynamics:

- *Synthetic population* data for various scenarios, data are generated by our Population dynamics model.
- *GIS data* for the location of the Staré Hradisko oppidum, i.e. maps with specifications of distance, slope and wetness indices. Data are used in cellular automata based model in NetLogo. The individual cells are allocated topographical (slope, wetness, hydrology), land use suitability (vegetation, soil productivity), and economic (distance from settlements, distance from water sources) variables. The system dynamics model in Stella does not operate with spatial data.
- *Diet specification* (e.g. the ratio of cereals, pulses, meat and milk) – there are numerous consequences. Higher cereal consumption requires more intensive growing with ploughing and manuring; more frequent manuring requires higher number of livestock, more animals require more working hours and the time consequently is missing in other activities including the intensive growing.
- *Strategy and work allocation settings* (e.g. intensive or extensive growing, workforce per 1 land unit and per 1 activity) – there are additional consequences and constraints. The total workforce has to be allocated according to the appropriate sequences of agricultural activities and their timing during the year (seeding, harrowing, ploughing, harvesting or manuring).

The model outputs are time series of land allocations (fields, pastures, woodland) and food production (cereals, pulses, meat, milk – all of them both consumed and stored to be used in years of the deficit). Similarly to the Population dynamics model, four scenarios were specified:

1. No event (*baseline*) – no unexpected events happen, i.e. the food production is not affected by inauspicious weather, fire, theft, animal decease etc. If the population grows, the food productions grew correspondingly up to the limits of the map (i.e. the carrying capacity).

2. Sudden event: *lost harvest* – the crop is damaged either by fire or weather (hailstorms, frosts, droughts etc.). The consumption requirements are not covered, the populations is expected to adapt to the situation.
3. Sudden event: *lost animals* – it can be caused by disease or theft. The meat and milk calories are not available and the oxen cannot be used for ploughing, therefore the arable land is affected and the agricultural strategy has to be modified.
4. Sudden event: *fire* – the crop in storage is destroyed (reserves from previous one or two years, seed for the next season), the populations is expected to adapt.

The first version of the model was implemented in the NetLogo. Only the cereal production was taken into account, as cereals covered the major part of the diet (about 75 percent). Thus the estimation of field area in relation to workforce gives us elementary notion of the carrying capacity of the location. Results were presented in [2]. In [6] we described the fuzzy plug-in for NetLogo and its application for evaluation of the land suitability.

The second version of the NetLogo model adds the pulses production and the livestock management (cattle, sheep, pigs and horses) and the detailed workforce management. The grid of patches is used for visualization of the map of the area: the settlement (oppidum) is surrounded by land of different types (fields, meadows and woodland). See Fig. 3 for two screenshots from the course of the simulation.

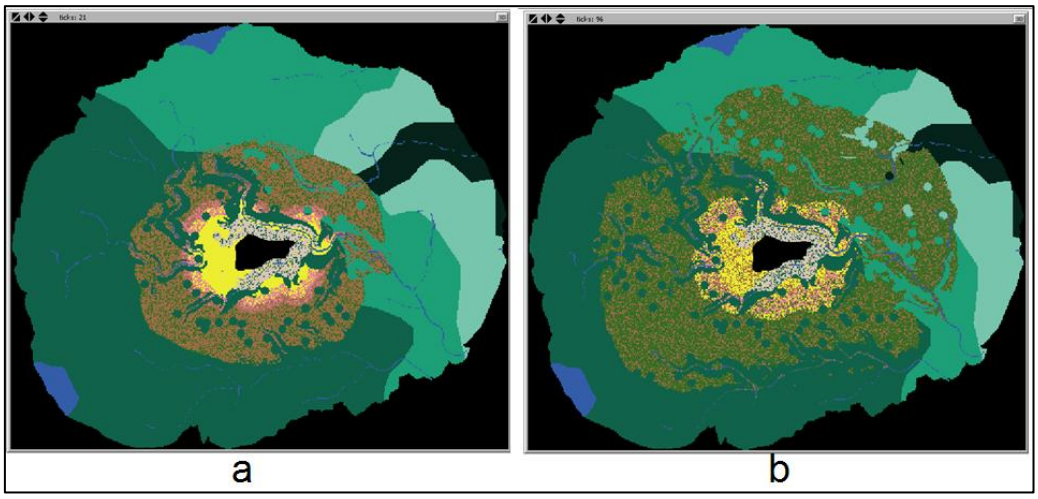


Figure 3 Land use model in NetLogo: (a) year 20, (b) year 90 Model showing the intensive (i.e. without fallow) agricultural strategy (yellow – cereals, pink – pulses, brown – pastures, i.e. deforested landscape). Output from the Population model, scenario with the gradual depopulation.

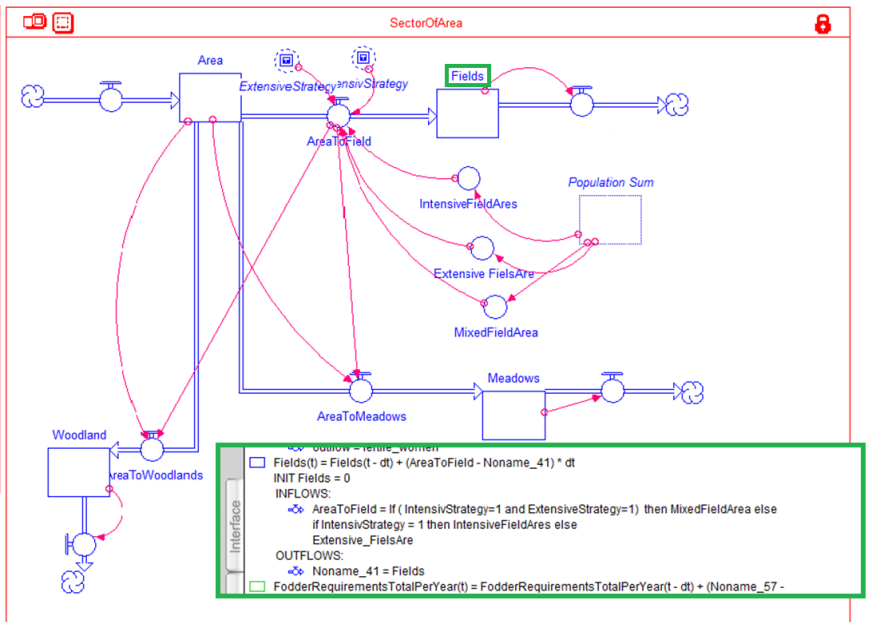


Figure 4 Land use model in Stella – part of the diagram, example of formula for Fields

The third version of the model is more abstract and aggregated. It was implemented in Stella. This system dynamics modelling tool does not allow us to visualize the map and observe changes in the land use. But on the other hand it simplifies the communication with domain experts (archaeologists), because the diagrams in Stella are well readable and the consistency of the model including its logic can be checked easily than in case of NetLogo models. Having two implementations, we can verify the independence of the simulation outputs on the implementation tool.

In Fig. 4, example of diagram is presented together with one of equations behind. The diagram defines the basic cycle of land use. The Area stock is connected with Fields stock, Woodland stock and Meadows (i.e. pastures) stock. The whole land use management is affected by the Population Sum stock, which has got its own diagram linked to the population dynamics model. Livestock management is defined analogically to the human population dynamics.

4 Conclusion

By analysing ancient economics we can effectively predict the outcomes for the modern one. Basic human requirements and means of their supplementation remain unchanged through history. Isolating and exploring the crisis factors can help us to understand the downfalls of the current economic development. Therefore good computational models are important for mastering the complexity of societies, for better understanding possible crisis factors and their relations to collapses of civilizations. We intend to provide explanatory models of functioning of Celtic society in Central Europe. The choice of appropriate modelling approach is limited by available data. There are no written sources providing deep details of everyday life of individuals. The knowledge of existence of some kind of nobility is not complete; the hierarchy of the society and the land allocation with respect to family ties is not proved. This lack of evidence does not allow us to specify daily routines of individuals and families, therefore the agent-based models similar to ENKIMDU ([1],[11]) cannot be designed. But more general models of carrying capacity of the settlements can be created. The system dynamics modelling is sufficient for designing aggregate models of population dynamics, land use and food production. The cellular automata represent a good alternative for capturing the changes of landscape correspondingly to human activities.

Our further objective is to finish descriptions of all models using ODD and ODD+D protocols ([4],[8]). In parallel we work on the model of interactions among the central places (oppida) and the villages in the hinterland. Villages with their individual characteristics will be represented by agents and organized in the network. This is the way how to simulate the supplier-customer relationships and to experiment with hypotheses about the level of self-sufficiency and the likely share of importing food and exporting the craft products.

Acknowledgements

The support of the Czech Science Foundation under Grant P405/12/0926 "*Social modelling as a tool for understanding Celtic society and cultural changes at the end of the Iron Age"* is kindly acknowledged.

References

[1] Christiansen, J.H., Altaweel, M.: Simulation of natural and social process interactions: An example from Bronze Age Mesopotamia. *Social Science Computer Review*, 2006, **24**(2), 1-17.

[2] Danielisová, A., Olševičová, K., Cimler, R., Machálek, T.: Understanding the Iron Age Economy: Sustainability of Agricultural Practices under Stable Population Growth. *Agent-based Modeling and Simulation in Archaeology* (Wurzer, G., Kowarik, K., Reschreiter, H., eds.), Springer, Advances in Geographic Information Science, Wien, 2013, 205-241.

[3] Danielisová, A.: „A pak se potichu vytratili…" Zánik keltské civilizace v Čechách z pohledu archeologického. *Kolaps a regenerace: cesty civilizací a kultur* (Bárta, M, Kovář, M. et al.), Academia, Praha, 2013, 107-138.

[4] Grimm, V. et al.: A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, 2006, **198**, 115-126.

[5] Heckbert, S.: MayaSim: An Agent-Based Model of the Ancient Maya Social-Ecological System. *Journal of Artificial Societies and Social Simulation*, 2013, **16** (4) 11.

[6] Machálek, T., Cimler, R. Olševičová, K., Danielisová, A.: Fuzzy Methods in Land Use Modeling for Archaeology, *Mathematical Methods in Economics*, 2013, Jihlava, 552-557.

[7] Machálek, T., Olševičová, K., Cimler, R. (2012) Modelling Population Dynamics for Archaeological Simulations. In: *Mathematical Methods in Economics*, 2012, Karviná, 536-539.

[8] Müller, B. et al.: Describing Human Decisions in Agent-based Models – ODD + D, An Extension of the ODD Protocol. *Environmental Modelling and Software,* 2013, **48**, 37-48.

[9] Olševičová, K., Cimler, R., Machálek, T.: Agent-based Model of Celtic Population Growth: NetLogo and Python. In: N.T. Nguyen et al. (Eds.) *Advanced Methods for Computational Collective Intelligence, Studies in Computational Intelligence 457,* Springer-Verlag Berlin Heidelberg, 2013, 135-143.
[10] Olševičová, K., Cimler, R., Tomášková, H., Danielisová, A.: Model of Carrying Capacity – System Dynamics in NetLogo and Stella. In: *Proc. of The International Conference Hradec Economic Days*, Hradec Králové, 2014, Part V, 117-123.
[11] Wilkinson, TJ., Christiansen, J.H., Ur, J., Altaweel, M.: Urbanization within a dynamic environment: Modeling bronze age communities in Upper Mesopotamia. *American anthropologist*, 2006, **109** (1), 52-68.

On the use of dispersion measures consistent with additive shifts

S. Ortobelli[1], T. Tichý [2], F. Petronio[3]

Abstract. In financial literature we can find many alternative dispersion and risk measures that can help us to identify a proper investments. However, only few dispersion measures are consistent with additive shifts. In this work we propose to study the class of dispersion measures that consider that "more is better than less" and we analyze and empirically compare their different characteristics. In particular, we discuss the property of consistency with respect to additive shifts and we examine a class of dispersion measures that satisfy this property. Finally, we examine the use of some of these dispersion measures in financial problems with particular attention to the portfolio selection problem.

Keywords: stochastic dominance, dispersion measure, portfolio selection.

JEL Classification: C58, G11
AMS Classification: 91B28

1 Introduction

In the portfolio choices, to identify a proper investment and discard the inefficient investment, some risk and/or dispersion measures must be adopted. In the last years, many dispersion and risk measures have been proposed in the literature. These risk measures are alternative to the classic standard deviation (variance) proposed by Markowitz and Tobin in the 50'. Among others we recall the recent classification of coherent measures of risk (Artzner et al. [1]), the mean-variance and mean-semivariance approaches (Markowitz [5]), mean absolute deviation (MAD) and semi-MAD approaches (see for example Konno and Yamazaki [4], Speranza [13], Ogryzak and Ruszczynski [6] or Zenios and Kang [14]), the mean-dispersion approach in the Dybvig's distributional analysis of portfolio choice (Dybvig [2]), and many others (see eg. Rachev et al. [10] and the references therein).

Markowitz and Tobin proposed a selection rule for a non-satiable risk-averse investor asserting that among a given set of investment alternatives, including the set of securities available at the market as well as all possible linear combinations of these basic securities, it is possible obtain the efficient set of alternatives by discarding those investments with a lower mean and higher variance than a member of the given set.

This selection rule is optimal for all non-satiable risk-averse investors under particular assumptions on return distribution or utility function, however it fails as an optimal general decision rule. In particular, as we see in Ortobelli, et al. [7] under the assumptions of "limited liability" and "no short sale", portfolios of gross returns are positive random variables and the mean-dispersion Markowitz and Tobin's rule is not optimal.

For this reason we are interested on the analysis of different dispersion measures defined for positive random variables, under some assumptions of adherence to the reality of the market. In particular, as suggested by Giacometti et al. [3] we are interested in dispersion measures that consider preferable random wealth W plus a positive constant a, than the only random wealth W.

The aim of this paper is to propose an empirical comparison among the standard deviation and the new measure consistent with additive shifts applied to different definition of returns that should be consistent with the investor preferences. For this purpose we consider an allocation problem under which the investors maximize different performance ratios based on different concepts of risk. Moreover, we consider a different definition of returns which we derive from recent returns classifications consistent with investor preferences, as shown in Ortobelli et al. [8].

The paper is organized as follows. In section 2 we recall the stochastic dominance characteristics of mean-dispersion approach and the properties of a recent proposed dispersion measures (see Giacometti et al. [3]). In

[1] University of Bergamo /Department MSIA, Via dei Caniana 2, 24127 Bergamo, Italy, sol.unibg.it
[2] VŠB-Technical University of Ostrava /Faculty of Economics, Department of Finance, Sokolská 33, 70121 Ostrava, Czech Republic, tomas.tichy@vsb.cz
[3] VŠB-Technical University of Ostrava /Faculty of Economics, Department of Finance, Sokolská 33, 70121 Ostrava, Czech Republic, filomena.petronio @vsb.cz

section 3 we compare the Sharpe ratio with a new dispersion measure applied to the returns defined for different investor's behavior. Conclusions are given in the last section.

2. The portfolio allocation problem with dispersion measures consistent with additive shifts

In this section we recall the main results show in Giacometti et al. [3] in which the admissible dispersion measures are studied for the portfolio selection problem and their consistency with the maximization of the expected utility. In this work the authors consider the portfolios gross returns, $x'z$, positive random variables defined on the probability space $(\Omega, \Im, P)$, where $x \in R_+^n$ is the vector of non-negative allocations among n risky limited liability investments with gross returns $z = [z_1,..., z_n]'$. Moreover, the authors assume that short sale is not allowed (i.e. $x_i \geq 0; \forall i$).

The last years have seen a heated debate on the risk measures. In particular, Artzner et al. [1] have provided a definition of coherent risk measures. A coherent risk measure applied to the portfolio selection problem with institutional restrictions is a non-positive real mapping ρ that associates a real number $\rho(x'z) \leq 0$ to each portfolio $x'z$. Such mapping has to satisfy the following axioms, in order to be a *coherent risk measure*:

1. For all portfolios $x'z$ and all real numbers a we have $\rho(x'z + az_0) = \rho(x'z) - a$, where z_0 is the riskless gross return.
2. For all portfolios $x'z$ and all $a \geq 0$ we have $\rho(ax'z) = a\rho(x'z)$.
3. For any $x'z$ and $y'z$, $\rho(x'z + y'z) \leq \rho(x'z) + \rho(y'z)$.
4. For any $x'z \leq y'z$, $\rho(x'z) \leq \rho(y'z)$.

Under these assumptions a typical coherent risk measure associated to the portfolio selection problem is $\bar{\rho}(x'z) = -\frac{E(x'z)}{z_0}$. As argued by Rachev et al. [10], part of the properties of a coherent risk measure can be subject to many criticisms. As observed by Giacometti at al. [3] could exist a dispersion measure, or increasing function of dispersion measures, which verify other not less interesting properties even if they do not support the four axioms of coherence. On the other hand, several dispersion measures and risk measures are still widely used in portfolio allocation problems (see Rachev et al. [10]). Recall that a dispersion measure is any law invariant measure σ (i.e., if X and Y have the same distributions, then $\sigma_X = \sigma_Y$) that satisfies the following characteristics:

1. (positive homogeneous) $\sigma_{\beta X} = \beta\sigma_X$ for every positive real $\beta \geq 0$;
2. (positive) $\sigma_X \geq 0$ and $\sigma_X = 0$ if and only if X is a constant a.s.;
3. (consistent with additive shift) $\sigma_{X+t} \leq \sigma_X$ for every real positive $t \geq 0$.

Most of the dispersion measures used in financial modeling are deviation measures. Therefore, by definition (see Rockafeller et al. [11]), all of them are the subadditive (i.e., $\sigma_{X+Y} \leq \sigma_X + \sigma_Y$ for any random variable X and Y) dispersion measures for which property 3 is satisfied as equality (i.e., $\sigma_{X+t} = \sigma_X$ for any real t).

As in Giacometti et al. [3], we assume any portfolio $x'z$ belonging to $L^q = \left\{X \mid \mathrm{E}(|X|^q) < \infty\right\}$ for some real q greater than or equal to one. Thus, if there exists finite $\mathrm{E}(\log(x'z))$,

$$\mathrm{cov}(x'z, \log x'z) = \mathrm{E}(x'z \log x'z) - \mathrm{E}(x'z)\mathrm{E}(\log x'z)$$

could be considered as dispersion measure because $\mathrm{cov}(x'z, \log x'z) < \infty$. Defining $\mathrm{colog}(X)$ the $\mathrm{cov}(X, \log X)$, we recall the main properties of this dispersion measure.

Proposition 1. *Assume that for each portfolio* $x'z$ *there exists finite* $\mathrm{E}(\log x'z)$ *and for some real* q *greater than or equal to one* $x'z \in L^q$. *Then:*

1) $\mathrm{colog}(x'z)$ *is a dispersion measure, according to the above definition. In particular, for every non constant positive random variable* X *and for each* $t>0$, $\mathrm{colog}(X+t)<\mathrm{colog}(X)$ *(i.e. it is strictly consistent with additive shifts).*

2) *For every couple of positive non constant random variables* X *and* Y*, if every risk-averse investor prefers* X *to* Y*, then* $\mathrm{colog}(Y)\geq\mathrm{colog}(X)$.

For a proof of this proposition see Giacometti et al. [3].

Moreover as a consequence of some results of ordering theory we can distinguish the definition of the returns with respect to the investors preferences as follows from the following lemma and definition (see Ortobelli et al. [8]):

Lemma 1 *Given two random variables X and Y , then the following equivalence relationships hold.*

1) (Martingale property) Every risk averse investor (i.e. with concave utility function) prefers X to Y if and only if there exist two random variables X', Y' defined on the same probability space $(\Omega,\Im,\mathrm{P})$ *that have the same distribution of X and Y such that:*

$$E(Y'|X')=X'\ a.s..$$

2) (Super-martingale property) Every non-satiable risk-averse investor (i.e. with increasing concave utility function) prefers X to Y if and only if there exist two random variables X', Y' defined on the same probability space $(\Omega,\Im,\mathrm{P})$*that have the same distribution of X and Y such that:*

$$E(Y'|X')\leq X'a.s..$$

3) (Sub-martingale property) Every non-satiable risk-seeking investor (i.e. with increasing convex utility function) prefers X to Y if and only if there exist two random variables X', Y' defined on the same probability space $(\Omega,\Im,\mathrm{P})$ *that have the same distribution of X and Y such that:*

$$E(X'|Y')\geq Y'a.s..$$

The proof of this lemma is a well-known result of ordering theory that can be found in Shaked and Shanthikumar [12].

Let $X=\{X_t\}_{t\in\mathbb{N}}$ be a discrete time wealth process and assume that the investor's temporal horizon is T. According to Rachev et al. [10], we generally use the average return (for unity of time of the process X)

$$r_s=\left(\frac{X_s}{X_{s-T}}\right)^{1/T}-1$$

in order to consider the aggregate risk of the period [s,s+T]. However, in view of Lemma 1, we can distinguish different possible return definitions with respect to the investors' preferences, the investors' temporal horizon and the portfolio wealth strategy.

Definition 1 *Let us assume that the wealth process of a given portfolio strategy follows a discrete stochastic process* $X=\{X_t\}_{t\in\mathbb{N}}$ *defined on a filtered probability space. Assume the investor has a temporal horizon equal to T (unities of time of the wealth process) and consider the real number* $\varepsilon\in[0,1]$. *Then for* $s>T$ *we call*

1) risk averse returns of the wealth process X for investors with temporal horizon T

$$r_s=\left(\frac{X_s}{E(X_{s-T}|X_s)}\right)^{\frac{1}{T}}-1;$$

2) risk seeking returns of the wealth process X for investors with temporal horizon T

$$r_s=\left(\frac{E(X_s|X_{s-T})}{X_{s-T}}\right)^{\frac{1}{T}}-1;$$

3) ε *-risk averse returns of the wealth process X for investors with temporal horizon T*

$$r_s=\left(\varepsilon\frac{X_s}{E(X_{s-T}|X_s)}+(1-\varepsilon)\frac{E(X_s|X_{s-T})}{X_{s-T}}\right)^{\frac{1}{T}}-1;$$

Observe that the returns r_s defined in (1), (2) and (3) represent average rates of return during the period [s, s+T]. Moreover, if we use the ε-risk averse returns of the wealth process we implicitly assume that the investors are

non-satiable neither risk averse nor risk seeking. The above returns can be easily determined using the conditional expected value estimator proposed by Petronio et al. [9].

3. Empirical analysis

In this section we propose an empirical comparison among different performance measure, applied to the returns defined for different investor behavior. For this purpose, we consider allocation problems where the investors maximize different performance ratios based on different concepts of risk. In particular, we consider the performance measures Sharpe ratio and Colog ratio given by $\frac{\mathrm{E}(x'z)-z_0}{\sigma_{x'z}}$, where $\sigma_{x'z}$ is the standard deviation or the Colog measure, respectively.

We consider the data from December 2003 to February 2014 of the S&P 100 components, taken from Thomson Reuters DataStream. We compare the wealth obtained maximizing different strategies: Colog mixed applied to ε-risk averse returns (with $\varepsilon = 0.5$), Colog risk-averse applied to risk averse returns, Colog risk-seeking applied to risk seeking returns, Colog classic, Sharpe ratio.

As showed in Figure 1, there are no significant differences between the strategies of different kind of investor except that the risk seeking type of investors will probably lose more wealth than others during this period of crisis. In fact we can observe significant losses in all scenarios considered. However, we can observe that the Colog mixed is a strategy more conservative than others for all period considered.

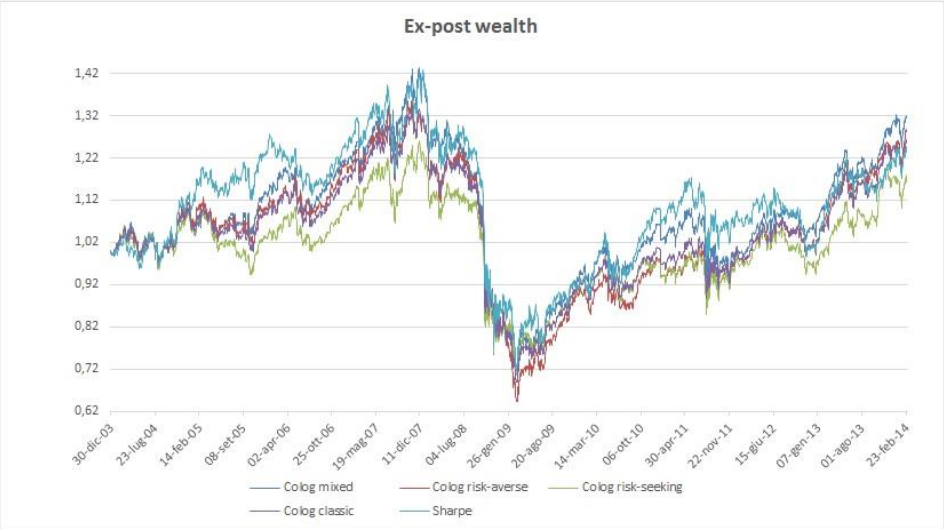


Figure 1 Evolution of wealth (ex-post), obtained maximizing different performance measures

The results observed in Figure 1 are confirmed in Table 1 where the main statistical indices involved in the data analysis are showed.

In particular, in Table 1 we summarized the average values of mean, standard deviation, skewness, kurtosis, and Sharpe value, to describe and compare, from a statistical point of view, the returns of the selected data samples, for the different strategies.

	Mean	St. dev.	Sharpe	Skewness	Kurtosis	Final wealth
Colog mixed	0.000158	0.009849	0.016011	-0.44956	8.85310611	1.321572
Colog risk-averse	0.000146	0.009798	0.014919	-0.17840	10.4865308	1.285133
Colog risk-seeking	0.000107	0.009537	0.011237	-0.24889	8.08742587	1.170638
Colog classic	0.000136	0.009523	0.014234	-0.31445	9.25006199	1.259114
Sharpe	0.000138	0.010566	0.013064	-0.30507	6.82628111	1.233560

Table 1 Descriptive statistics for particular strategies

4. Conclusion

In this work we proposed a study of the class of dispersion measures that consider that "more is better than less" and we analyzed and empirically compared their different characteristics. In particular, we recall the property of consistency with respect to additive shifts and a class of dispersion measures that satisfy this property. We use these measures in financial problems with particular attention to the portfolio selection, comparing the dispersion measure applied to the returns defined for different investor behavior. In our analysis the optimization of the Colog mixed performance measure presents good results in terms of final wealth in particular during the period of crisis.

Acknowledgements

The research was supported through the Czech Science Foundation (GAČR) under project 13-13142S and SP2014/16, an SGS research project of VSB-TU Ostrava, and furthermore by the European Regional Development Fund in the IT4Innovations Centre of Excellence project CZ.1.05/1.1.00/02.0070 (second author), including the access to the supercomputing capacity, and the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296 (first author) and CZ.1.07/2.3.00/30.0016 (third author).

References

[1] Artzner, P., Delbaen, F., Heath, J-M., Eber, D.: Coherent Measures of Risk. *Mathematical Finance* **9** (2000), 203–228.
[2] Dybvig, P.: Distributional analysis of portfolio choice. *Journal of Business* **61** (1988), 369–393.
[3] Giacometti, R., Ortobelli, S., Tichy, T.: Portfolio selection with uncertainty measures consistent with additive shifts, *Prague Economics Papers* (to appear in 2014).
[4] Konno, H., Yamazaki, H.: Mean-absolute deviation portfolio optimization model and its application to Tokyo stock market. *Management Science* **37** (1991), 519–531.
[5] Markowitz, H.: *Portfolio selection; efficient diversification of investment*. Wiley, New York, 1959.
[6] Ogryczak, W., Ruszczynski, A.: On consistency of stochastic dominance and mean-semideviation models. *Mathematic Programing* **B89** (2001), 217–232.
[7] Ortobelli S. Rachev S., Shalit H., Fabozzi F.: Orderings and probability functionals consistent with preferences. *Applied Mathematical Finance* **16**(1) (2009), 81–102.
[8] Ortobelli, S., Petronio, F., Lando, T.: Portfolio problems based on returns consistent with the investor's preferences, in (J. Balicki ed.) Advanced in Applied and pure mathematics, Proceeding of MCSS14, 2014, 340-347.
[9] Petronio, F., Lando, T., Biglova, A., Ortobelli, S.: Optimal portfolio performance with Exchange traded funds. *Ekonomická revue – Central European Review of Economic Issues*, **17** (2014), 5–12.
[10] Rachev, S., Ortobelli, S., Stoyanov, S., Fabozzi, F., Biglova, A.: Desirable properties of an ideal risk measure in portfolio theory. *International Journal of Theoretical and Applied Finance*, **11** (2008), 19–54.
[11] Rockafellar, R. T., Uryasev, S., Zabarankin, M.: Generalized Deviations in Risk Analysis. *Finance and Stochastics*, **10** (2006), 51–74.
[12] Shaked, M. and Shanthikumar, J. G., *Stochastic Orders and Their Applications*, Academic Press, San Diego, 1994.
[13] Speranza, M. G.: Linear programming models for portfolio optimization. *Finance*, **14** (1993), 107–123.
[14] Zenios, S. A., and Kang, P.: Mean-absolute deviation portfolio optimization for mortgage-backed securities. *Annals of Operations Research*, **45** (1993), 433–450.

Correlation analysis of regional disparities – the unemployment and crime in the regions of the Czech Republic

Zuzana Palová[1], Milena Botlíková[2], Josef Botlík[3]

Abstract. Although the decrease in criminality has been in recent years, the crimes brutality is increasing year by year. Unemployment is often mentioned as one of the causes of these crimes. Unemployment often leads to a deterioration of the social status of the unemployed. This can subsequently lead to the increase of crime from minor violations and thefts to serious bodily harms. The paper is devoted to the correlation analysis and data modeling in SPSS program of regional disparities – unemployment rate and the number of crimes per 1,000 inhabitants. The analysis will be performed on the level of the regions of the Czech Republic NUTS 3 for the years 2005 - 2012. The aim of this paper is to prove or disprove a direct relation between unemployment and increase of criminality. Our hypothesis presumes that the increase of the unemployment rate will be followed by the increase in the number of crimes committed. The correlation coefficients for each region to confirm or refute our hypothesis will be the result of this work. Based on the identified dependencies, analysis of crime development and its dependency on the growth or decline of crime rate will be performed.

Keywords: correlation analysis, regional disparities, regions of the Czech Republic, unemployment rate, number of crimes per 1 000 inhabitants, modeling data

JEL Classification: C22, E24, O18
AMS Classification: 62H20, 62M10

1 Introduction

Although the decrease has been in criminality in recent years, the crimes brutality is increasing year by year. It is often stated that one of the causes of criminality is increasing unemployment rate. This paper tries to verify that it is indeed so. Stated hypothesis presumes that the increase of the unemployment rate will be followed by the increase in the number of crimes per 1,000 inhabitants. Based on the identified dependencies, analysis of development of criminality and its dependency on the growth or decline of criminality will be performed.

The unemployment rate and the number of crimes per 1,000 inhabitants belong to the two most significant regional disparities. This paper is focused on the influence of the unemployment rate on the criminality in the Moravian-Silesian Region (MSR) and what the influence of the unemployment rate in the MSR has on the criminality in other regions. The data for the analysis were used for the years 2005 – 2012 (source: Czech Statistical Office).

2 Regional disparities

By using regional disparities it is possible to trace the regional disparities, identify, and then to structure measure and evaluate them. Regional disparities can be divided into three spheres - social, economic and territorial. These spheres are further broken down into problem areas, descriptors and finally indicators (Fachineli, Tomanek, 2011).

For this analysis, we selected two indicators - the unemployment rate (*UNE*) and the number of crimes per 1,000 inhabitants (*CRIME*). These are indicators of two different spheres respectively spheres of economic and social. Because it is often assumed that the individual indicators across spheres are related, the individual indicators could be measured between themselves.

[1] Silesian University, School of Business Administration, Univerzitní náměstí, 1934/3, 733 40 Karviná, zuzana.palova@centrum.cz.

[2] Silesian University, School of Business Administration, Univerzitní náměstí, 1934/3, 733 40 Karviná, botlikova@opf.cz.

[3] Silesian University, School of Business Administration, Univerzitní náměstí, 1934/3, 733 40 Karviná, botlik@opf.cz.

3 Analysis crimes and unemployment

3.1 Crimes per 1 000 inhabitants

Criminality in the Czech Republic has a decreasing tendency almost in all regions of the Czech Republic. Only in two regions increase in crime occurred in the period in the Vysočina Region (increase 1) and in the Moravian-Silesian Region (increase 5). The largest decrease in crime during the period was recorded in Prague, where there was even a drop by 23.3 crimes per 1 000 inhabitants.

In 2012 (see Fig. 1) it was recorded most of the crimes per 1 000 inhabitants in Prague (58.1), this region was followed by the Karlovy Vary Region (33.2), followed by the Moravian-Silesian Region (33.1). In contrast, the regions with the lowest number of crimes per 1 000 inhabitants were Zlín Region (15.1), Vysočina Region (16.7) and the Pardubice Region (17.4).

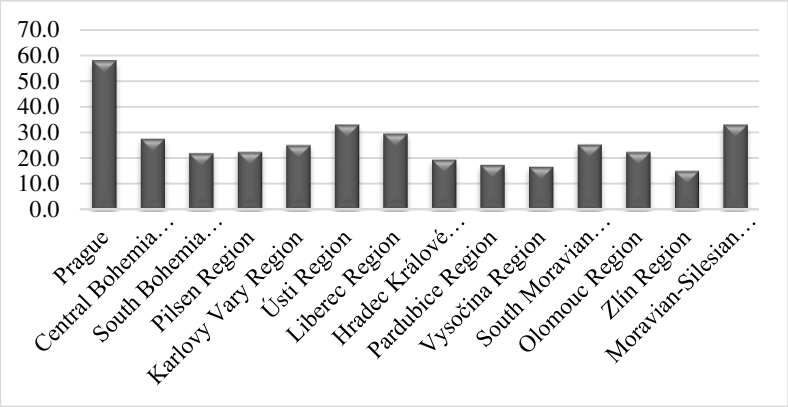


Figure 1 Crimes per 1 000 inhabitants (in 2012)

3.2 Unemployment rate

As oppose to the indicator of crime, the unemployment rate doesn't tend to decrease. In the reporting period there was a decline in the unemployment rate in only five regions. The highest decrease was recorded in the Moravian-Silesian and Ústí Region. All regions in which there was a decline in the unemployment rate (excluding the Pardubice Region - the decline was only slight) were among the regions with the highest unemployment rate in the Czech Republic. At the beginning of the period the difference between the regions with the lowest and highest unemployment rates was 12 percentage points. In 2012 compared to the beginning of the period there was a decrease disparities in the labor market and to 9.1 percentage points.

Figure 2 shows the unemployment rate in each region in 2012. The chart shows that the highest unemployment rate was in the Ústí, Olomouc and Moravian-Silesia Region. In contrast, the regions with the lowest unemployment rates were Prague, Pilsen and the Central Bohemia Region.

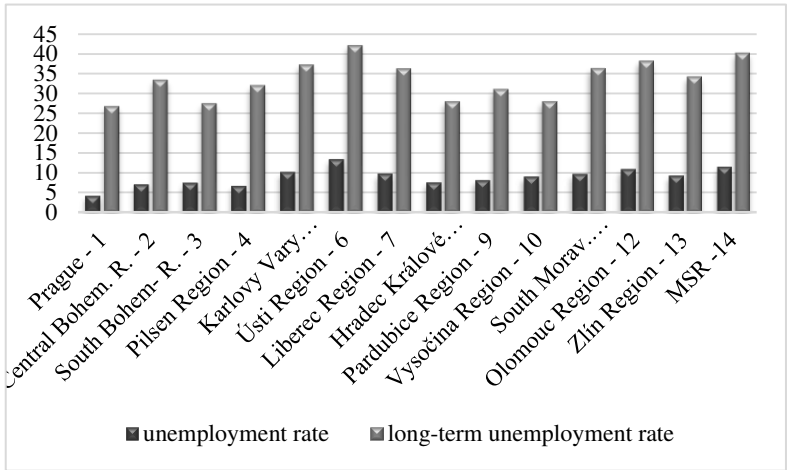


Figure 2 The average unemployment and long-term unemployment rate in % (in 2012)

From an overall assessment of the indicators we can see that the Moravian-Silesian and the Ústí Region are the most affected regions of crime (excluding the city of Prague). These regions also belong to the regions with

the highest unemployment rate in the CR. These regions in the period managed to reduce the level of unemployment. The MSR but compared to the Usti Region couldn't reduce in criminality. In the MSR and the Vysočina Region there was in 2012 compared to 2005 an increase in criminality (the highest increase among regions).

3.3 Pearson correlation analysis

The Spearman's coefficient or the Pearson correlation coefficient can be used for the analysis of the closeness of relationships between individual characteristics (indicators of disparities). The correlation coefficient determines the closeness of relationships, mostly linear. On the basis of the correlation coefficient it is not possible to determine the direction of the relationship. This means that an increase or decrease in unemployment leads to an increase/decrease in criminality or, conversely, increase/decrease in criminality will be accelerator of growth or decline in unemployment, i.e., it does not address the causality, i.e., causes and consequences.

The Pearson's selective coefficient (1) was used for the analysis of the closeness of relationships. This coefficient is suitable for data with the normal distribution, which the file shows. [4] The value of the coefficient r can take values in the range (-1; 1).[5] If the values of the coefficient are positive, there is an increase/decrease of one variable (X) to the growth of second variable (Y). If the correlation coefficient takes negative values, the individual characteristics have inverse relationship (with the increase/decrease (X) there is a decrease/increase the second (Y)). Chraska (2003) considers for the optimal value of the coefficient ± 0.45.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

		UNE1	UNE2	UNE3	UNE4	UNE5	UNE6	UNE7	UNE8	UNE9	UNE10	UNE11	UNE12	UNE13	UNE14	LUNE6	LUNE10	LUNE14
CRIME2	Pearson Correlation	**-0,92**	**-0,89**	**-0,87**	**-0,73**	**-0,75**	-0,43	**-0,74**	**-0,80**	**-0,76**	**-0,81**	**-0,73**	**-0,79**	**-0,77**	-0,29	0,11	-0,18	0,19
	Sig. (2-tailed)	0,00	0,00	0,00	0,04	0,03	0,28	0,04	0,02	0,03	0,02	0,04	0,02	0,03	0,49	0,79	0,68	0,65
CRIME3	Pearson Correlation	-0,47	-0,41	-0,52	-0,46	-0,30	0,30	-0,69	-0,31	-0,31	-0,50	-0,14	-0,30	-0,35	0,42	0,71	0,57	**0,79**
	Sig. (2-tailed)	0,24	0,31	0,19	0,25	0,47	0,48	0,06	0,45	0,45	0,20	0,74	0,47	0,40	0,30	0,05	0,14	0,02
CRIME4	Pearson Correlation	**-0,76**	**-0,71**	**-0,76**	-0,65	-0,57	-0,08	**-0,77**	-0,61	-0,57	-0,70	-0,48	-0,58	-0,61	0,08	0,42	0,17	0,50
	Sig. (2-tailed)	0,03	0,05	0,03	0,08	0,14	0,84	0,02	0,11	0,14	0,05	0,23	0,13	0,11	0,85	0,30	0,70	0,20
CRIME6	Pearson Correlation	**-0,75**	**-0,73**	**-0,79**	**-0,74**	-0,63	-0,06	**-0,88**	-0,66	-0,67	**-0,78**	-0,52	-0,65	-0,68	0,08	0,62	0,37	0,70
	Sig. (2-tailed)	0,03	0,04	0,02	0,04	0,09	0,88	0,00	0,08	0,07	0,02	0,19	0,08	0,06	0,85	0,10	0,37	0,06
CRIME7	Pearson Correlation	**-0,89**	**-0,85**	**-0,86**	**-0,75**	-0,69	-0,19	**-0,84**	**-0,73**	**-0,71**	**-0,81**	-0,61	**-0,72**	**-0,73**	-0,03	0,37	0,11	0,45
	Sig. (2-tailed)	0,00	0,01	0,01	0,03	0,06	0,64	0,01	0,04	0,05	0,01	0,10	0,04	0,04	0,93	0,36	0,79	0,26
CRIME8	Pearson Correlation	**-0,82**	**-0,85**	**-0,86**	**-0,81**	**-0,79**	-0,39	**-0,82**	**-0,80**	**-0,81**	**-0,84**	**-0,72**	**-0,82**	**-0,81**	-0,27	0,33	0,00	0,39
	Sig. (2-tailed)	0,01	0,01	0,01	0,02	0,02	0,34	0,01	0,02	0,02	0,01	0,04	0,01	0,01	0,52	0,43	0,99	0,34
CRIME9	Pearson Correlation	**-0,72**	**-0,72**	**-0,78**	**-0,76**	-0,63	-0,04	**-0,88**	**-0,63**	**-0,66**	**-0,77**	-0,50	-0,63	-0,69	0,10	0,59	0,34	0,67
	Sig. (2-tailed)	0,05	0,04	0,02	0,03	0,09	0,93	0,00	0,09	0,08	0,03	0,20	0,09	0,06	0,81	0,12	0,41	0,07
CRIME11	Pearson Correlation	**-0,80**	**-0,80**	**-0,82**	**-0,80**	-0,69	-0,11	**-0,89**	-0,70	**-0,74**	**-0,83**	-0,58	**-0,73**	**-0,74**	0,01	0,57	0,33	0,63
	Sig. (2-tailed)	0,02	0,02	0,01	0,02	0,06	0,79	0,00	0,05	0,03	0,01	0,13	0,04	0,04	0,97	0,14	0,43	0,10
CRIME12	Pearson Correlation	-0,55	-0,63	-0,67	**-0,73**	-0,66	-0,28	-0,72	**-0,64**	-0,69	-0,69	-0,58	-0,65	-0,70	-0,20	0,41	0,12	0,44
	Sig. (2-tailed)	0,16	0,09	0,07	0,04	0,07	0,49	0,04	0,09	0,06	0,06	0,13	0,08	0,05	0,64	0,31	0,78	0,27
CRIME13	Pearson Correlation	**-0,90**	**-0,86**	**-0,85**	**-0,72**	-0,70	-0,29	**-0,77**	**-0,74**	**-0,71**	**-0,79**	-0,65	**-0,74**	**-0,73**	-0,14	0,24	-0,04	0,31
	Sig. (2-tailed)	0,00	0,01	0,01	0,04	0,05	0,49	0,03	0,03	0,05	0,02	0,08	0,04	0,04	0,75	0,57	0,93	0,45
CRIME14	Pearson Correlation	0,16	0,10	0,16	0,14	-0,09	-0,66	0,38	-0,10	-0,05	0,15	-0,24	-0,07	-0,03	**-0,75**	**-0,71**	**-0,71**	**-0,73**
	Sig. (2-tailed)	0,71	0,82	0,70	0,75	0,84	0,07	0,36	0,82	0,90	0,73	0,56	0,87	0,94	0,03	0,05	0,05	0,04

Figure 3 Pearson correlation matrix

Determining of the significance of the coefficient at the selected level α is the condition for confirming the tightness relationship (in the most cases a linear dependency). The determination of significance, whether it is true: H_0: $\rho = 0$ the independence of two variables compared to the alternative hypothesis $H1$: $\rho \neq 0$ the linear dependence of variables, can be done by comparing the theoretical value of T (2) a critical value t (2).

[4] Test of data apportionment - test

[5] The values close to 0 indicate the non-correlation and the values approaching to 1of the tightness of data (correlation 1about tightness).

$$T = r\sqrt{\frac{n-2}{1-r^2}} \qquad > \qquad t \Rightarrow |t|_{(n-2;\,1-\alpha/2)} \tag{2}$$

Where: n number o parameters, α level of significance.

It is recommended to confirmed the statistical significance of the correlation coefficient on the basis of measured of the correlation coefficient with r^2 (coefficient of determination). In case that $r^2 > 0.1$ H_0 does not apply and the existence of a dependency between the different characteristics (indicators) assuming statistical significance is confirmed. For determining statistical significance of the correlation coefficient within the investigated sample there will be used the results generated from SPSS (from the Sig. - significance). The acceptance of $H1 \neq 0$ of statistical significance *r* and thus the relationship between individual characteristics it is determined by the relationship $Sig. < 0.05$ (α - level of significance 0.05).

As is evident from the correlation matrix the unemployment in different regions has a direct effect (linear relationship; gray box) on crime in each region. The presumption that the crime will increase with the increase of unemployment or opposite relationship (with the increase of criminality the unemployment will increase) was confirmed only in the case of a linear relationship of long-term unemployment[6] of MSR with the criminality of South Bohemian Region. Other statistically significant correlation coefficients show an inverse relationship (with increasing indicator is reduced the second indicator).

3.4 Model - development of the criminality

Regression analysis allows finding of the causes and consequences. The influence of unemployment on criminality in the MSR and the effect of unemployment on criminality in other regions will examine from the chart's expression of selected linearly dependent variables. From the chart 4b) it is possible to assume that there is a linear function between long-term unemployment in the MSR and the development of criminality in the South Bohemian Region, i.e. an increase of criminality in South Bohemia Region is caused by long-term unemployment in the MSR (Figure 4), which it cannot be said for monitored values from the first chart (linear regression function may have errors).

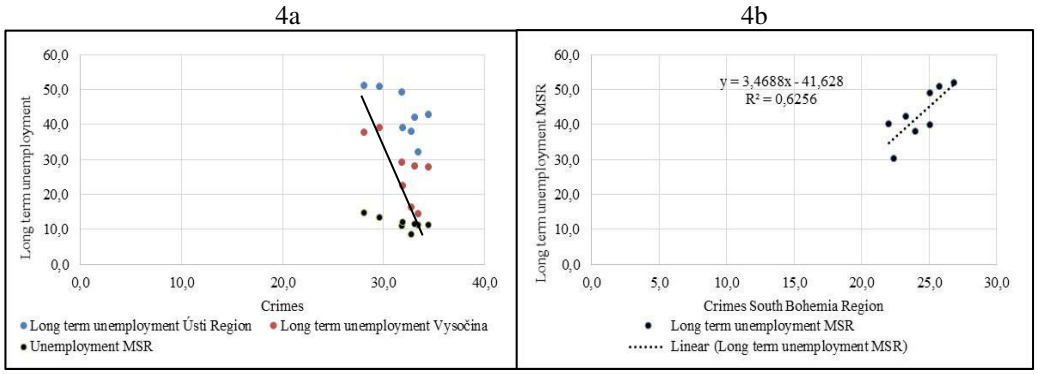


Figure 4 Graphic representation

Specification of independent variables (unemployment and long-term unemployment) and the dependent variable (crimes) which explain consequence of the general regression model (3) was a precondition for the creation the model that captures the relationship between unemployment and criminality (3).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots \beta_n x_n + \varepsilon \tag{3}$$

The following model (4) was created by generating the dependent variable (*criminality in the MSR*) and independent variables *long-term unemployment in the Ústí (6), Vysočina (10) and Moravian-Silesian Region (14b , and unemployment in the MSR (14a*) through a program SPSS.

$$\begin{aligned} y_{crimi} &= 48.9 - 1.21\,unemplymen\,t\ MSR + 0.59\,long\ tern\ unemployme\,nt\,Ústí \\ &+ 0.19\,long\,tern\,unemployme\,nt - 0.79\,long\,tern\ \ unemployme\,nt\,MSR \end{aligned} \tag{4}$$

[6] Long-term unemployment rate: share of long term unemployed of total labor force in %.

R	R^2	F	Sig. F	Sig. β_1 (14a)	Sig. β_2 (6)	Sig. β_3 (10)	Sig. β_4 (14b)	DW
0.932	0.868	4.931	0.11	0.094	0.426	0.488	0.190	2.54

Table 1 Evaluation of the quality of model

Condition of quality testing is required for the adoption of the model. Quality of the model (see Table 1) is also evaluated based on the determination coefficient[7]. Although coefficient of determination is sufficiently high ($R^2 = 0.868$ close to 1), estimates of individual parameters exhibit the statistical insignificance (evaluating the size Sig, significance is the case $Sig.\ \beta_n < 0.05$). There was also performed *F-test* for determining the significance of the model itself and comparing the test criterion and critical value[8]. The test criterion *(T)* is less than the critical value $F_{k,n-k-1}(\alpha)$, which may be indicative of a poorly chosen model (5). The model shows considerable multicolinearity (see Table 2). It is difficult to decide about the autocorrelation *Durbin – Watson*[9].

$$T = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}} (4.93) < (9.12) F_{k;n-k-1}(\alpha) \tag{5}$$

	Unemployment MSR	Long term unemployment Ústí	Long term unemployment Vysočina	Long unemployment MSR
Unemployment MSR	1	0.565	0.632	0.801
	.	0.145	0.093	0.017
Long term unemployment Ústí	0.565	1	0.988	0.900
	0.145	.	0.000	0.002
Long term unemployment Vysočina	0.632	**0.988**	1	0.940
	0.093	0.000	.	0.001
Long unemployment MSR	**0.801**	**0.900**	**0.940**	1
	0.017	0.002	0.001	.
Crimes MSR	**-0.754**	**-0.730**	**-0.712**	**-0.713**
	0.031	0.040	0.047	0.047

Table 2 Correlation matrix

Based on the poor quality of the model, the individual regressors (characteristics) were gradually excluded so that the model was not affected by errors in the model (multicollinearity, autocorrelation, significance) and model was created revealing the relationship between criminality and long-term unemployment in the MSR (6).The model was tested on the significance of the model and estimating the parameters. The significance values indicate the statistical significance of the model ($F\ (6.862) > F_{(0,05;\ 1;6)}(5.59);\ Sig.\ \beta_1 < 0.05$).

$$y_{crimi} = 40.76 - 0.21\ long\ termunemployment\ MSR \tag{6}$$

R	R^2	Adjusted R Square	Change Statistics				
			R Square Change	F change	df1	df2	Sig.F Change
0.730	0.534	0.456	1.552	**6.862**	1	6	0.047

	Unstandardized Coefficients	Standardized Coefficients	t	Sig.		
	B	Std. Error	Beta			Low Bound	Upper Bond
Constanta	40.762	3.427		11.894	0,000	32.376	49.147
Long term unemployment MSR	-0.207	0.079	-0,730	-2.620	**0.040**	-0.401	-0.014

Figure 5 Summary of model 1

[7] Coefficient of determination force-fit of the data, R^2 values in the interval <0; 1>, values close to 1 indicate the snugness.

[8] Multicollinearity represents the presence of strongly correlated variables in the model (variables evolving in the same direction). It may be indicative of a wrong model. Revealed of the multicollinearity can be done on the basis of the correlation matrix if any of the correlation coefficient is above 0.8 the addiction is harmful.

[9] DW optimal value = (1.5;2). DW (2,54) is in the interval ($4-d_u$ $4-d_L$)

The value of *DW* is *1.602* and simultaneously DW > d_U, 1.602 > 1.332 (table values for the test Durbin - Watson), H_1 with the presence of autocorrelation is rejected. (H_0 with absence of autocorrelation is received).

As follows from the Pearson correlation matrix (see Table 3) the linear relationship was demonstrated between long-term unemployment in the MSR and criminality in the South Bohemia Region and long-term unemployment in the MSR. In this case, there is by increasing of one indicator the increase of the second indicator. Based on the assessment of model's quality (R^2 *=0.63; Sig.* β_1 *< 0.05; F (10.19) >* $F_{(k.\ n-2)}(\alpha)$ *(5.99)* it is possible to adopted the model. Transformation of the data was used to remove autocorrelation (7). SPSS model of long-term unemployment in the MSR and criminality in the South Bohemia Region was generated from the adjusted data (8;9).

$$y_{crimi} = 16.54 - 0.21\, long\ term unemployment\ MSR \tag{7}$$

$$r = \frac{\sum_{i=2} e_{i-1} e_i}{\sum_{i=1} e_t} = 0.19 \tag{8}$$

$$y_1^* = y_1 \sqrt{(1 - r^2)} \quad x_1^* = x_1 \sqrt{(1 - r^2)}$$
$$y_n^* = y_n - y_{n-1} * r \qquad x_n^* = x_n - x_{n-1} * r \tag{9}$$

R	**R^2**	**Adjusted R Square**	**Change Statistics**				
			R Square Change	F change	df1	df2	Sig.F Change
0.859	0.738	0.695	1.548	**16.935**	1	6	0.007

	Unstandardized Coefficients	**Standardized Coefficients**	**t**	**Sig.**		
	B	Std. Error	Beta		Low Bound	Upper Bond
Constanta	10.005	2.512		3.983	0,007	3.858	16.153
Long term Unemployment MSR	0.286	0.069	0.895	4.115	0.006	0.116	0.455

Table 3 Model of regression analysis

4 Conclusion

The regression model shows that there is an inverse relationship between long-term unemployment in the MSR and criminality in the MSR. The model shows that an increase of 1 % in long-term unemployment in the MSR leads to a decrease of 0.207 crimes per 1,000 inhabitants in the MSR. Furthermore, it was found if there is an increase of 1 % in long-term unemployment in the MSR then there will be an increase of 0.29 per 1,000 inhabitants in the South Bohemian Region. The hypothesis of unidirectional linear dependence between the studied social and economic indicators was not confirmed from the correlation analysis (with the exception of the criminality in the South Bohemian Region and long-term unemployment in the MSR). Base on this analysis, it could be assumed the existence of interregional migration of the unemployed.

Acknowledgements

This paper originated from the project SGS 20/2014 Analysis of the business environment in the region of Karviná.

References

[1] Bilingsley, P.: *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.
[2] Dupačová, J., and Römisch, W.: Quantitative stability for scenario-based stochastic programs. In: *Proceedings of the Prague Stochastics '98* (Hušková, M., Lachout, P., and Víšek, J. Á., eds.). Union of Czech Mathematicians and Physicists, Prague, 1998, 119–124.
[3] Chraska,, M.: Uvod do výzkumu v pedagogice. (Základy kvantitativně orientovaného výzkumu Univerzita Palackého. (2003)
[4] Fachinelli, H., Tománek P.: *Hodnocení regionálních disparit prostřednictvím integrovaných indikátorů*. Regionální disparity. Elektronický časopis. 2011, 31-44.
[5] Ramík, J.: Statistika (pro magisterské stadium. (2007) OPF Karviná
[6] Tošenovský, F.: Ekonometrické metody. OPF v Karviné, skriptum, (2013) s. 61
[7] Vallander, S. S.: Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability and its Applications* **18** (1973), 784–786.

A Heuristic Approach to the p-Median Problem Using a Set Covering Formulation

Stanislav Palúch [1], Ivana Urbaničová [2]

Abstract. This paper presents an attempt to solve the p-median problem in a two step procedure. The first step creates a set of all nearest customers with feasible capacity for every possible depot. The second step consists in choosing an optimum subset of the family of sets created in the first step such that every customer is covered at least by one chosen set.

Keywords: Capacitated p-median problem, Set covering problem, Heuristic algorithm

JEL classification: C61
AMS classification: 90C06, 90C10, 90C27, 90C59

1 Introduction

There are many objectives when we solve the problem to decide where to build a facility. Objectives for location of coal depots are different from objectives for rescue ambulances and both mentioned are different from ones for deciding where to locate waste disposal sites. An attempt for classification of location problems was made by Blaško in [1]. A proposal for location emergency ambulances in region of Žilina is in [3].

However, the largest attention is paid to effort to locate facilities in order to minimize transportation costs. The last mentioned problem is known as the p-median problem. If we suppose that the facilities do not have limited capacities we have an (ordinary or non capacitated) p-median problems. If facilities have limited capacities we have capacitated p-median problem CPMP.

Capacitated p-median problem is NP-hard and has wide ranging applications in designing transportation and communication networks, in designing public service systems, emergency services etc. Over the past decades, the problem has received much attention in solution methods. Several algorithms including exact, heuristic and metaheuristic methods have been proposed to solve this problem. A good survey of methods and corresponding references is contained in [8]. A new approach to facility location appeared in last years – namely to take into account not only economical point of view but also so called fairness – a parameter which requires that there are no serviced places with significantly worse service than average standard.

2 Mathematical model of capacitated p-median

Suppose we are given an edge weighted graph $G = (V, E, c)$ where $V = \{1, 2, \ldots, n\}$ and where $c(e)$ is the length of the edge $e \in E$. Let d_{ij} be the distance between two vertices $i,\ j$. (Distance matrix $\{d_{ij}\}_{i=1,2,\ldots,n,\ j=1,2,\ldots,n}$ can be easily calculated using e.g. Floyd algorithm.) Let $F \subseteq V$ be the set of vertices where facilities can be located. Let $C \subseteq V$ be the set of customers who are to be served by a facility. We will suppose that $C = V$. Denote by q_i the demand of each customer $i \in V$ and denote by Q_j the capacity of the possible facility located in vertex $j \in F$. Let x_{ij} $i \in V$, $j \in F$, y_j, $J \in F$ are binary decision variables with the following meaning:

[1] University of Žilina /Faculty of Management Science and Informatics, Department of Mathematical Methods, Univerzitná 8215/1 010 26 Žilina, Slovakia, stanislav.paluch@fri.uniza.sk

[2] University of Žilina /Faculty of Management Science and Informatics, Department of Mathematical Methods, Univerzitná 8215/1 010 26 Žilina, Slovakia, ivana.urbanicova@fri.uniza.sk

$$x_{ij} = \begin{cases} 1 & \text{if customer } i \in V \text{ is served by facility located in } j \in F \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$y_j = \begin{cases} 1 & \text{if a facility is located in vertex } j \in F \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Suppose that a facility is located in $j \in F$. Then $y_j = 1$ and total expenses of this facility can be calculated as $\sum_{i \in V} q_i d_{ij} x_{ij}$. Therefore total expenses for all located facilities equal $\sum_{j \in F} \sum_{i \in V} q_i d_{ij} x_{ij}$.

Capacitated p-median problem can be formulated as follows:

$$\text{Minimize} \quad \sum_{j \in F} \sum_{i \in V} q_i d_{ij} x_{ij} \quad (3)$$

$$\text{subject to} \quad \sum_{j \in F} x_{ij} = 1 \quad \text{for all } i \in V \quad (4)$$

$$\sum_{j \in F} y_j = p \quad (5)$$

$$\sum_{i \in V} q_i x_{ij} \leq Q_j y_j \quad \text{for all } j \in F \quad (6)$$

$$x_{ij},\ y_j \in \{0, 1\} \quad \text{for all } i \in V \text{ and for all } j \in F \quad (7)$$

Constraints (4) ensure that every customer is assigned to exactly one facility, constraint (5) guarantees that exactly p facilities are allocated and conditions (6) make sure that capacities of all allocated facilities are not exceeded.

3 An attempt to solve CPMP as a set covering problem

The result of optimization problem (3) – (7) is a partitioning of the set V into p districts, every one of which is characterized by its serving facility j and contains all vertices $i \in V$ such that $x_{ij} = 1$. Moreover, the total demand of every such district $\sum_{i \in V} q_i x_{ij}$ is less or equal than the capacity Q_j of its serving facility.

The idea of our approach is first to define a district for every candidate vertex $j \in F$ where a facility can be located. In the second step it is necessary to calculate the cost for every such district. The third step consists in choosing p districts covering all vertices from V with the minimum total cost. The final step is to solve situation where some vertices appear in more than one chosen district. This can be done simply by assigning such vertex to nearest facility.

Let $D_j = \{v_{j,1}, v_{j,2}, \ldots, v_{jk_j}\}$ be a predetermined district of candidate $j \in F$. District D_j can be created by setting $D_j = \{j\}$ and then by stepwise adding nearest vertex v from $V - D_j$ such that capacity of $D_j \cup \{j\}$ is less of equal to Q_j. Let $C(D_j)$ be the cost of the district D_j. In our studied case $C(D_j) = \sum_{v \in D_j} q_v d_{jv}$.

Now we can define constants b_{ij}:

$$b_{ij} = \begin{cases} 1 & \text{if } v_i \in D_j \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Define a decision variable z_j

$$z_j = \begin{cases} 1 & \text{if facility located in } j \text{ is selected} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

The problem to cover every vertex $v \in V$ by at least one district D_j with the minimum total cost can be formulated as follows:

$$\text{Minimize} \quad \sum_{j \in F} C_j z_j \tag{10}$$

$$\text{subject to} \quad \sum_{j \in F} b_{ij} z_j \geq 1 \quad \text{for all } i \in V \tag{11}$$

$$\sum_{j \in F} z_j = p \tag{12}$$

$$z_j \in \{0,1\} \tag{13}$$

$$\tag{14}$$

Ideal situation would occur if all selected districts were disjunctive ones. Unfortunately, this is not the case. Therefore the final solution is obtained by assigning every vertex which is contained in more than one district to the nearest facility.

One way how to reduce the number of vertices in overlapping selected districts is the following procedure. We create a sequence $D_j^1, D_j^2, \ldots, D_j^k$ of nested districts $D_j^1 \subset D_j^2 \subset, \ldots, \subset D_j^k = D_j$ (instead of single D_j) for every candidate vertex $j \in F$ with increasing capacities.

Let $N = |F|.k$ Let $\overline{D}_1, \overline{D}_2, \ldots, \overline{D}_N$ be an enumeration of all districts from $\{D_j^p, \mid j \in F,\ p = 1, 2, \ldots, k\}$, let $\overline{c}_1, \overline{c}_2, \ldots, \overline{c}_N$ are corresponding costs of districts. Denote by $\overline{b}_{ij}$ for $i \in V$ and $j = 1, 2 \ldots, N$ a constant and denote by z_{ij} a binary decision variable with the following meaning:

$$\overline{b}_{ij} = \begin{cases} 1 & \text{if } v_i \in \overline{D}_j \\ 0 & \text{otherwise} \end{cases} \qquad z_j = \begin{cases} 1 & \text{if district } \overline{D}_j \text{ is selected} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

$$\text{Minimize} \quad \sum_{j=1}^{N} \overline{C}_j z_j \tag{16}$$

$$\text{subject to} \quad \sum_{j=1}^{N} \overline{b}_{ij} z_j \geq 1 \quad \text{for all } i \in V \tag{17}$$

$$\sum_{j=1}^{N} z_j = p \tag{18}$$

$$z_j \in \{0,1\} \tag{19}$$

One could rise a question whether the optimum solution $z_i,\ i = 1, 2, \ldots, N$ of (16)–(19) can contain two districts for the same j.

Suppose that there are two $\overline{D}_i$, $\overline{D}_k$ two districts adherent to the same j. In this case $\overline{D}_i \subset \overline{D}_k$ or $\overline{D}_k \subset \overline{D}_i$. If $\overline{D}_i \subset \overline{D}_k$ (i.e. if $z_i = 1$ and $z_k = 1$) then the solution defined by $z'_j = z_j$ for all $j \neq i$ and $z'_i = 0$ is a feasible solution of (16)–(19) with lesser value of objective function what is in contradiction with optimality of z_i.

4 Computational results

We used data of Slovak road network adapted for rescue ambulances. This network contains 2921 vertices representing communities as customers. Number of inhabitants were given for every community – this number was considered as the demand q_i of corresponding vertex $i \in V$. We used precomputed distance matrix since this problem does not need detailed structure of the network. We considered all communities as possible candidates for facility location, i.e. $F = V$. Capacities of vertices were not available, therefore we decided to use uniform capacity Q the same for all vertices. This capacity has to be greater or equal then the maximum of demands of all customers – $Q \geq \max\{\} q_{ij}$, otherwise no feasible solution does exist. The cost of district D_j was calculated as $\sum_{i \in D_j} q_i d_{ij}$.

Task	Used districts with capacity k*Q											XPRESS				Percentage of	
	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	Processing time [s]	p	Number of chosen depots	Value of objective function	People in more than one depot	cities in more than one depot
SR_1	1											7	273	55	3442	40	44
SR_2	1	1										14	273	56	3213	34	42
SR_3	1		1									21	273	61	2771	35	42
SR_4	1			1								27	273	70	2479	36	43
SR_5	1				1							28	273	78	2220	35	45
SR_6	1					1						26	273	92	1995	33	43
SR_7	1						1					23	273	117	1637	38	47
SR_8	1							1				34	273	154	1244	36	49
SR_9	1								1			33	273	176	1402	29	44
SR_10	1									1		39	273	220	1882	24	40
SR_11	1										1	56	273	260	2242	22	37
SR_30	1	1	1	1		1						260	273	81	2256	28	37
SR_31	1	1	1	1			1					233	273	98	1753	30	40
SR_32	1	1	1	1				1				267	273	140	1268	32	45
SR_33	1	1	1	1					1			230	273	191	1038	30	45
SR_34	1	1	1	1						1		231	273	242	1390	22	41
SR_35	1	1	1	1							1	327	273	273	1799	15	34
SR_54	1	1	1	1	1	1	1	1	1	1		38	273	273	517	18	37
SR_55	1	1	1	1	1	1	1	1	1		1	68	273	273	587	13	30
SR_56	1	1	1	1	1	1	1	1	1	1	1	148	273	273	509	12	30
SR_65	1		1		1		1				1	19	273	273	988	16	36
SR_66	1		1		1		1		1		1	31	273	273	603	14	32

Table 1 Dependance of objective function value and overlapping percentage on the set of nested districts

First experiment was done using model (3)–(7) for $p = 273$ District D_j was constructed by adding nearest customers until capacity Q was overrun. The result was astonishing – enormous percentage of customers was assigned to more that one facility. This was caused by existence several vertices with $q_j > 150000$ and 273*150000=40 million – Slovakia should have at least 40 million inhabitants if the chosen districts would be disjoint.

Therefore the condition (5) was changed to $\sum_{j \in F} y_j \leq p$. Overlays have diminished but still remained too large. The reason is that customers demands range from tens to hundred thousands and therefore uniform facility capacity is extremely inconvenient for regions with small demand.

For that reason we decided to use model (16)–(19) with (18) changed to $\sum_{j=1}^{N} z_j \leq p$ with capacities $1Q,\ 0.9Q,\ \ldots,\ 0.2Q,\ 0.1Q,\ 0.05Q$.

5 Conclusion

One of the advantages of proposed set covering approach is the possibility of using more general objectives than till now used transportation costs provided that the generalized cost is an ascending function, i.e. if $D_i \subset D_k$ then $c(D_i) < c(D_k)$. Another benefit is its applicability to very large instances.

A drawback of our approach is that quite large percentage of vertices is contained in more than one chosen depot in our computed instance. Maybe this was caused by extremal inhomogeneity in demand of individual vertices and by unavailability of their real capacities. Therefore we had to use an uniform capacity equal to the maximum capacities of all customers. (If the uniform capacity is smaller then the maximum of demands, the corresponding capacitated p-median problem had no solution.)

Our experiments showed that todays computers can deal with large instances in reasonable time. Contrary to our instance where all vertices were also facility candidates in real word problems the set F of candidates is substantially lesser then the set V of all vertices. This fact offers the following way of improvement of our approach.

We used a set of nested districts $\{D_j^1 \subset D_j^2 \subset, \ldots, \subset D_j^k = D_j\}$ for a candidate j. Instead of a set of nested sets it is possible to take into account more general set of districts $\{D_j^1, D_j^2, \ldots, D_j^k = D_j\}$ with the property that $c(D_j^p) + c(D_j^q) < \min\{c(D_j^p), c(D_j^q)\}$ for $1 \leq p, q \leq k$. Further research will be focused on smaller sets of candidates F with given capacity Q_j for every $J \in F$. Initial computation will start with a nested set of districts for every candidate j. Afterwards the set of districts for every j can be modified and set covering procedure can be applied with modified sets of districts.

Acknowledgements

This work was supported by the research grant APVV-0760-11 Designing of Fair Service Systems on Transportation Networks

References

[1] Blaško, R.: Lokačné úlohy a ich stručná klasifikácia (Location problems and their concise classification), *Horizonty dopravy*, **1** (2004), VUD Žilina, ISSN 1210-0978

[2] Ceselli, A.: Two exact algorithms for the capacitated p-median problem. *Journal of Operations Research* (2003) **4** pp. 319-340

[3] Gábrišová, L., Janáček, J.: Návrh rozmiestnenia staníc záchrannej služby v Žilinskom kraji. *Sborník příspěvku Optimalizace pokrytí území veřejnými obslužnými systémy,* Univerzita Pardubice, (2013), pp.: 16-25

[4] Ghoseiri, K., Ghannadpour, S.,F.: Solving Capacitated P-Median Problem using Genetic Algorithm, (2007) IEEE 1-4244-1529-2/07/

[5] Janáček, J.: *Optimalizace na dopravních sítích*, EDIS - vydavateľstvo Žilinskej univerzity, (2002), ISBN 80-8070-031-1

[6] Lorena, L.,A.,N., Senne, E.,L.,F.: A column generation approach to capacitated p-median problems. *Comput Oper Res* (2004) 31, pp.: 863–876

[7] Mladenovic, N., Hansen, P., Moreno-Prez, J.,A.: The p-median problem: a survey of metaheuristic approaches. *European Journal of Operations Research* (2007) 179, pp.: 927–939

[8] Yaghini, M., Momeni, M., Sarmadi, M., H, R,: An efficient heuristic algorithm for the capacitated p-median problem, *Operations Research* Springer-Verlag Berlin Heidelberg (2013) **11** 229–248.

Capital mobility measurements

Václava Pánková[1]

Abstract. Capital as one of the most important productive inputs is - under globalization - characterized by a certain mobility degree, which can be a significant factor that influences economic growth. According to the Feldstein – Horioka hypothesis, perfect capital mobility means that domestic savings and investment rates are uncorrelated. The optimality of capital mobility can be considered by using the consumption-smoothing current account approach: the mobility of capital flows is optimal when it allows for full consumption smoothing over time through the dynamics of the current account in the face of shocks to domestic output, investment, and government expenditure. The consumption-smoothing approach is based on comparing consumption-smoothing part of current account and optimal current account constructed by the help of expected increments of net foreign assets constructed on the base of relevant VAR model formulation. An application to the Czech Republic and Slovakia is performed finding that none of both countries exhibit an optimal level of capital mobility.

Key words: capital mobility, current account, VAR model, Granger causality.

JEL Classification: C01, C12, E22
AMS Classification: 91G70

1 Introduction

Capital mobility is defined as the ability of private funds to move across national boundaries in the pursuit of the efficient allocation of resources and higher returns. Usually, it increases national savings and investment and impacts on capital accumulation and economic growth. If capital is mobile, then it is easier to attract FDI into a country; also investment opportunities abroad increase. Technology transfers, production chains and trade and financial development improve efficiency. However, a positive influence of an open capital account emerges only after a country has achieved a certain degree of economic development. Common experience shows that capital mobility is appreciated in different ways according to the economic development of a country. Though some literature (e.g. [8]), describes international capital mobility as today's economic orthodoxy, it is necessary to admit that a sudden outflow of capital in times of uncertainty may occur.

To measure the degree of capital mobility, different approaches appear in the literature. In [3] and [11], two alternative measures of the extent of capital mobility are presented and explained, both constructed as indices during the 1990s.

An econometric explanation of capital movements was influenced by Feldstein – Horioka [4] hypothesis: If capital is perfectly mobile, domestic savings and investment should be weakly correlated, hence, studying a possible cointegration relation an answer to capital (non)mobility question can be found. Relevant parameter being estimated quantifies the measure of capital mobility in given economy but it did not answer a question about a suitability of the result for the given economy.

Recent studies (e.g. [7]) propose to explain international capital mobility using a consumption – smoothing current account. The basic idea states (see e.g. [7]) that without impediments to the movement of capital the current account behaves like a cumulative power in the face of shocks to the key economic variables as output, investment and government spending to ensure a constant level of consumption over time. The level of mobility of capital is

[1]University of Economics, Dept. of Econometrics, Winston Churchill sq. 4, 130 67 Prague, Czech Republic
e-mail: pankova@vse.cz

understood as an optimal one, if it allows for full consumption – smoothing over time through the dynamics of the current account.

Macroeconomic adjustment to changes in economic environment are to a large extent conditioned by the inter-temporal decisions of economic agents. Based on expectations of future dynamics of key variables, they alter current indebtedness what implies changes in future consumption possibilities. J. Sachs [10] thus argues that current account as a measure of changes in national net indebtedness, depends as much on future economic trends as on the current economic environment. He also formulates a model of the current account departing from inter-temporal optimization problem. The integral part of the model is a consumption – smoothing motive which corresponds with the permanent income hypothesis.

2 Theoretical background

Current account *CA* is defined as the change in the value of a country´s net foreign assets. *CA* is also identically equal to the capita account balance including the change in reserves, assuming the balance of payments accounting.

$$CA_t = Y_t - G_t - I_t + rB_t - C_t \tag{1}$$

where *Y* is gross domestic output, *G* is government spending, *I* investment, *C* aggregate consumption, *B* foreign assets and *r* is the world interest rate. Besides,

$$CA_t^{CS} = Y_t - G_t - I_t + rB_t - \theta C_t \tag{2}$$

represents the consumption – smoothing part of *CA* (see Sachs, 1982) where θ is a parameter that removes the trend in consumption. If *CA* is I(0), CA^{CS} also is I(0), hence $Y - G - I - rB$ and *C* are cointegrated what offers a way how to find parameter θ.

In economics, the concept of net foreign assets (*NFA*) relates to balance of payments identity, only the traditional balance of payments identity does not take into account changes in asset prices and exchange rates. *NFA* also known as net output is defined by

$$NFA_t = Y_t - G_t - I_t \tag{3}$$

and relates to optimal CA^* (see e.g. [5]) by

$$CA_t^* = -\sum_{s}^{\infty}\left(\frac{1}{1+r}\right)^s E_t(\Delta NFA_{t+s}|R_t)\,. \tag{4}$$

In (4), *R* is the current information set. The representative agent will decrease its current account by means of borrowing in world capital markets if a future increase in income is expected, and vice versa.

Future values of ΔNFA can be found by the help of the VAR model formulated as

$$\begin{bmatrix} \Delta NFA_t \\ CA_t \end{bmatrix} = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \begin{bmatrix} \Delta NFA_{t-1} \\ CA_{t-1} \end{bmatrix} + \begin{bmatrix} \mu_{\Delta NFA} \\ \mu_{CA^{CS}} \end{bmatrix} + \begin{bmatrix} u_{t1} \\ u_{t2} \end{bmatrix}. \tag{5}$$

For the following considerations, the VAR model should be stationary.

If capital mobility in the given economy is optimal, (2) and (4) values should be identical (see e.g. [1]). The eventual equality is studied by analyzing relevant restrictions in the VAR model according to the following process.

Rewriting (5) in the matrix form we have

$$X_t - \mu = \Psi(X_{t-1} - \mu) + u_t \tag{6}$$

and for CA^* constructed according to (4) it is

$$CA_t^* - \mu_{CA^*} = K(X_t - \mu) \tag{7}$$

with K being a vector

$$K = \begin{bmatrix} 1 & 0 \end{bmatrix} \frac{1}{1+r} \Psi (I - \frac{1}{1+r} \Psi)^{-1}. \tag{8}$$

Testing

$$H_0 : CA_t^* - \mu_{CA^*} = CA_t^{CS} - \mu_{CA^{CS}}$$

we ask about the validity of

$$CA_t^* - \mu_{CA^*} = CA_t^{CS} - \mu_{CA^{CS}} = \begin{bmatrix} 0 & 1 \end{bmatrix} (X_t - \mu). \tag{9}$$

Combining (7), (8) and (9) the question changes into a checking of restrictions on the elements of matrix Ψ. For not rejecting the null we should find

$$\psi_{11} = \psi_{21}, \ \ \psi_{12} = \psi_{22} \text{ and } \psi_{12} = \psi_{22} - (1+r)$$

according to Wald test.

As it is derived in [5], $K = \begin{bmatrix} \alpha & \beta \end{bmatrix}$ with

$$\alpha = \frac{1}{\gamma} \left[-\psi_{11}(1 + r - \psi_{22}) - \psi_{12}\psi_{21} \right]$$

$$\beta = \frac{1}{\gamma} \left[-\psi_{11}\psi_{12} - \psi_{12}(1 + r - \psi_{11}) \right]$$

$$\gamma = 1 + 2r - \psi_{22} + r^2 - r\psi_{22} - \psi_{11} - \psi_{11}r + \psi_{11}\psi_{22} - \psi_{12}\psi_{21}.$$

The null hypothesis corresponds with $\alpha = 0, \ \ \beta = 1$.

The key assumption evidently is CA^{CS} Granger – cause ΔNFA. Otherwise, $\psi_{12} = 0$ hence, $\beta = 0$ and we reject the null.

To summarize the procedure we have

(a) Verifying stationarity of CA_t, ΔNFA_t.
(b) Calculating parameter θ.
(c) Checking if CA^{CS} Granger – cause ΔNFA.
(d) Testing of restrictions given on parameters.

3 Application

The procedure described above occurs in different modifications in recent literature. In [6] the factors of income level and/or geographic region are stressed. She concludes that the consumption - smoothing behavior is more often evident in advanced economies. In [2] the authors distinguish according to natural resource richness. Economies of G7 are analyzed Gaglione and Issler in [5] with a finding that the model is valid for Canada but the null has to be rejected for U.S.A. and Japan what contradicts previous findings. Adler [1] who deals with the economy of Sweden observed that there are many cases when the model fails.

Being interested in the economies of Visegrád, the author remains her own computations (to be published), that the capital immobility persists in Poland; owing to non-significant regression results, there is no clear answer in

the case of Hungary. Evident capital mobility is confirmed in the Czech Republic and Slovakia. That is why the above model will be applied to CR and Slovakia only. Using the annual data covering 1996 to 2013 (source Eurostat) and the method described above, an eventual optimality of capital mobility in CR and SR is studied.

The results are summarized in Table 1.

	Czech Republic	Slovakia
CA stationary *	yes	yes
ΔNCF stationary **	yes	yes
parameter θ ***	1.521	2.049
CA^{CS} Granger- cause ΔNFA	no	no
Testing of restrictions	not relevant	not relevant

Table 1 Results – own computations

* Having small dimension samples, KPSS test of stationarity was applied not rejecting null (CA is stationary) at 1 % level of significance.
** KPSS, 10 % level of significance
*** Despite of theoretical conclusions, in both cases cointegration was not traced between Y – G – I – rB and C by applying Johansen test. As it may be caused by an unsufficient dimension of data, both entities were regressed and residuals tested for unit root by KPSS with stationarity confirmed (CR – 5 %, SR – 10 %). Parameter θ results from the regression. Author is aware of inaccuracy of such a technique.

Estimating the VAR models, Granger non - causality was apparent. The above statement concerning rejection of the null by $\beta = 0$ takes place.

4 Conclusions

Application of the model shows that in the Czech Republic as well as in Slovakia, the existing capital mobility does not equal its optimal level. The quality of the results is unfavourably influenced by two facts: First, not very long period during which both economies participate in the global capital movements. Second, model proposed is closely connected with the permanent income hypothesis (PIH) factor. As the author documents in [9], both countries exhibit a consumption according to PIH at a 50 % level only. Future studies are intended to comprise the excess volatility effect according to Flavin [5] which can be interpreted as relaxing the assumption of PIH and also of perfect international capital mobility.

Acknowledgements

Financial support of VŠE IGA F4/24/2014 is gratefully acknowledged.

References:

[1] Adler, J.: The open economy excess sensitivity hypothesis: Theory and Swedish evidence, Working Papers in Economics No. 88, 2002, https://gupea.ub.gu.se/bitstream/2077/2816/1/gunwpe0088.pdf
[2] Arajuo, J., Li, B., Poplawski-Ribeiro, M., Zanna L.F.: Current Account Norms in Natural Resource Rich and Capital Scarce Economies, IMF Working Paper, 2013, http://www.imf.org/external/pubs/ft/wp/2013/wp1380.pdf
[3] Edwards, S.: Capital Mobility and Economic Performance: Are Emerging Economies Different?, NBER Working Paper No. w.8076, 2001 http://www.cedeplar.ufmg.br/economia/disciplinas/ecn933a/crocco/Liberaliza caoregulamentacao_contas_capitais/ EDWAR~18.PDF
[4] Feldstein, M.; Horioka, C.: Domestic Saving and International Capital Flows, *Economic Journal* **90** (358), pp. 314–329, 1980

[5] Flavin, M. A.: The Adjustment of Consumption to Changing Expectations about Future Income, *The Journal of Political Economy*, Vol. 89, No. 5, pp. 974-1009, 1981
[6] Gaglianone, W.P., and Issler, J.V.: An Econometric Contribution to the Intertemporal Approach of the Current Account, Working Paper 178, Banco Central do Brasil, 2008, http://www.bcb.gov.br/pec/wps/ingl/wps178.pdf
[7] Rakovski, T.: Consumption Smoothing Current Account and International Capital Mobility–Evidence from Time Series and Panel Data, McGill University 2009 http://digitool.library.mcgill.ca/webclient/ StreamGate? folder_id=0&dvs= 1385703336096 ~ 593
[8] Palley, T.: Rethinking the economics of capital mobility and capital controls, Brazilian Journal of Political Economy, vol. 29, no. 3 pp. 15–34, 2009
[9] Pánková, V.: Hypotéza permanentního příjmu v zemích visegrádské skupiny (PIH in V4), *E + M*, after reviewing process, accepted to be published
[10] Sachs, J.: The Current Account in the Macroeconomic Adjustment Process, *J. of Economics*, vol. 84, No. 2, pp 147 – 159, 1982
[11] Schneider, B.: Saving. Investment Correlation and Capital Mobility in Developing Countries, Working

Labour market rigidities: A DSGE approach

Adam Pápai[1], Daniel Němec[2]

Abstract. DSGE models are frequently applied tools in modern macroeconomic modelling. They are used to describe the behavior of the economy. The goal of this paper is to capture the dynamics of the labour market in a small open economy. For our estimation we chose a specific DSGE model and applied it to the Hungarian economy. This model contains search and matching mechanisms, while the wages and the worked hours are the result of the Nash bargaining problem. The wage and price rigidities are modelled according to the approach developed by Rotemberg. This method somewhat differs from the widely used Calvo price setting. The model is able to capture the great degree of openness of the Hungarian economy and the relatively high levels of frictions. Furthermore the model confirms the weakness of the trade unions. Surprisingly, the results show quite low influence of the labour market on the dynamics of the rest of the economy.

Keywords: DSGE model, labour market, frictions, search and matching, small open economy.

JEL classification: E32, J60
AMS classification: 91B40

1 Introduction

In this paper we try to investigate the labour market dynamics of the Hungarian economy in the time period from the second quarter of 2001 to the fourth quarter of 2013. Our selection therefore covers only the inflation targeting era. For the estimation, we use a dynamic stochastic general equilibrium (DSGE) model developed and estimated by Albertini et al. [1]. They used New Zealand data and managed to get a reasonable estimate of the behavior of the real economy.

Significant amount of papers with similar aiming were created in the past two decades. The article of Mortensen and Pissarides [7] can be considered pioneering in the developing of the search and matching framework. Their work serves as a basis for many researchers and macroeconomic modellers. Incorporating these search and matching mechanisms into their paper, Jakab and Kónya [4] estimated a model for the Hungarian economy for the period from 1995 to 2008. They divided the data to two periods given the introduction of inflation targeting in 2001 and carried out two separate estimations. They found out that many of the model parameters were affected only partially by this change, for example price stickiness slightly increased while the bargaining power of workers remained unaltered. The same model was also estimated for the Czech economy by Němec[9].

The price and wage frictions in this paper are modelled according to the approach developed by Rotemberg [12]. Same pricing method was chosen by Zanetti [13], who investigated the impact of introducing trade unions into a DSGE model. He found the model with unions to be better at replicating the business cycle properties of the euro area, than the model without them. Also using Rotemberg-style price stickiness, Neugebauer and Wesselbaum [8] compared models with and without wage and price rigidities. Their analysis on the United States data shows, that the model with rigidities is significantly better at matching the properties of the observed data, than the model with no frictions. Modelling nominal rigidities based on Rotemberg [12] differs in some aspects from the widely used Calvo [3] approach. As pointed out by Lombardo and Vestin [5], Rotemberg-style price adjustment causes real costs for all firms. On the other side, Calvo-style price setting allows only a section of firms to change their prices. This

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, papai@mail.muni.cz
[2]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, nemecd@econ.muni.cz

causes a movement of demand between firms with and without optimized prices. For detailed discussion about the two types of price setting see Ascari et al. [2], Lombardo and Vestin [5] or Nisticó [10].

1.1 Characteristics of Hungarian economy

The growth of Hungarian gross domestic product was really stable until the first signs of crisis appeared in 2007. The attempt to recover from this situation failed completely after the main wave of the financial crisis hit the country in 2008. Since then the output has a downward trend. In 2001 inflation targeting was applied by the central bank. The target was continuously adjusted down from 7% to its present value of 3%. This helped to reduce the inflation from over 10% in 2000 to 3% in 2013. The average unemployment rate in the examined period was around 8%. It grew gradually from 5.5% in 2002 to its maximum of more than 11% in 2010, partially because of the crisis. The worked hours were fluctuating almost perfectly around their average, thus did not change significantly in the long run. On the other hand, the wage was strongly influenced by the crisis and dropped considerably. The data show, that while wages and hours are less volatile than the output, the volatilities of unemployment and vacancies are much higher. Furthermore, the hours and the wages are lagged behind the output by three periods and the unemployment by one.

2 The model

The small open economy model developed by Albertini et al. [1] is estimated in this paper. It is characterized by a non-Walrasian labour marker with search and matching fictions based on Mortensen and Pissarides [7]. Homogenous workers and firms meet at this market and negotiate the wage and the amount of hours worked through a Nash bargaining process. No capital or government is present. Price and wage setting problem is separated between two distinct types of domestic firms. Foreign products are imported and sold unaltered on the final good market. The model contains several kinds of frictions. First, the matching function ensures a certain delay, so the unemployed do not find jobs instantly. Second, vacancy creation is costly, so the firms do not post jobs recklessly. Finally, firms have to pay a certain adjustment cost if they want to change the wages or prices.

2.1 Labour market

At each time period t, workers can be employed or unemployed. On the other side, job positions can be filled or vacant. The number of new worker–vacancy pairs, M_t is given by Cobb-Douglas matching function with constant returns to scale:

$$M_t = \varepsilon_t^m S_t^\nu V_t^{1-\nu}, \tag{1}$$

where ε_t^m is the shock of matching efficiency, S_t is the number of job seekers, V_t is the number of vacancies and parameter $0 < \nu < 1$ represents the elasticity of the matching function. The evolution of employment, N_t can be written as:

$$N_t = (1 - \rho^x)N_{t-1} + M_t, \tag{2}$$

where ρ^x is the exogenous and constant rate of job destruction. Considering constant labour force, L (and normalizing it to 1), the number of seekers is defined by the following equation:

$$S_t = 1 - (1 - \rho^x)N_{t-1}. \tag{3}$$

Therefore the number of seekers is different from the number of unemployed ($U_t = 1 - N_t = S_t - M_t$). Moreover, employees who lose their job in time t can be re-employed in the same period.

2.2 Agents

The model consists of three agents: homogenous households, firms, and monetary authority. The utility of households depends on the amount of leisure and consumption, C_t. The set of consumption goods is

defined as a combination of domestic and imported goods. The decision making is given by the following utility function:

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \varepsilon_t^c \left[\log(C_t - \vartheta \bar{C}_{t-1}) - N_t \kappa_t^h \frac{h_t^{1+\phi_h}}{1+\phi_h} \right]. \tag{4}$$

Parameter $0 < \beta < 1$ is the standard discount factor, ε_t^c is the preference shock, ϑ is the parameter of habit. κ_t^h represents the parameter of disutility from working, ϕ_h is the inverse of Frisch elasticity and h_t is the hours the household decides to work. The income of households consists of wage, unemployment benefits, interests from holding bonds and profits from holding firms shares.

Three kinds of firms are present in the economy. The first kind of the firms is represented by producers of domestic intermediate goods, who are the only firms hiring workers. They sell their products at a perfectly competitive market to the second type of firms, the domestic final good producers. Importers represent the third category. They offer differentiated foreign goods at the domestic market. The last two operate on monopolistically competitive markets, thus they can set the prices of their goods. However, they face quadratic price adjustment costs, as in Rotemberg [12].

Labour is the only input used by the intermediate good producers. The output, $Y_{I,t}$ is determined by a production function:

$$Y_{I,t} = \varepsilon_t^a (N_t h_t)^{\zeta}, \tag{5}$$

where ε_t^a is the productivity shock, ζ is the labor share, N_t represents the workers and h_t the hours. Maximization of profit comes from the following function:

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{\lambda_0} \left[mc_t Y_{I,t} - \frac{W_t}{P_{H,t}} h_t N_t - \Gamma(V_t) - \Upsilon(W_t) N_t \right], \tag{6}$$

where λ is the Lagrange multiplier from the households maximization problem. mc_t is the price for which the intermediate good producers sell their goods to the final good producers. W_t is the nominal wage, $\Gamma(V_t)$ is the hiring cost and $\Upsilon(W_t)$ is the wage adjustment cost. The hiring cost has the following form (taken from Rotemberg [11]):

$$\Gamma(V_t) = \frac{\kappa_t^v}{e} V_t^e, \tag{7}$$

where $\kappa_t^v > 0$ is the scaling factor and $e > 0$ is the elasticity of creating a vacancy. For the case when $e = 1$, the marginal cost of hiring a new employee is constant. The wage adjustment cost function has the following form:

$$\Upsilon(W_t) = \frac{\psi_W}{2} \left(\frac{\pi_{w,t}}{\tilde{\pi}_{w,t-1}} - 1 \right)^2 h_t Y_{I,t}, \tag{8}$$

where ψ_W is the wage adjustment cost parameter, $\pi_{w,t} = W_t / W_{t-1}$ is the wage inflation and $\tilde{\pi}_{w,t-1} = \pi_{w,t-1}^{\gamma_W} \bar{\pi}_w^{\gamma_W}$. Parameter γ_W serves as the weight of the wage from the previous period in the wage setting process. Wage is set along with the worked hours in a Nash bargaining process. Here the total surplus is being divided between the firms and the workers:

$$\max \varphi_t^{1-\xi_t} \mu_t^{\xi_t}, \tag{9}$$

where φ_t is the surplus of households, μ_t is the surplus of firms and parameter $0 < \xi_t < 1$ represents the bargaining power of firms.

The profit functions of the domestic final good producers and the importers look similar as the profit function in equation (6). However, they do not have wage and hiring costs and instead of wage setting cost, they face price adjustment costs.

Finally, the monetary policy is present via a simple Taylor rule:

$$i_t = i_{t-1}^{\rho_r} \left[\frac{1}{\beta} \left(\frac{E_t \pi_{t+1}}{\pi} \right)^{\rho_\pi} \left(\frac{Y_t}{Y} \right)^{\rho_Y} \left(\frac{Y_t}{Y_{t-1}} \right)^{\rho_{\Delta Y}} \left(\frac{e_t}{e_{t-1}} \right)^{\rho_e} \right]^{1-\rho_r} \varepsilon_t^m. \tag{10}$$

Thus when the monetary authority sets the nominal interest rates, it takes into consideration the interest rates from the previous period, the expected inflation, the output and the nominal exchange rate, e_t.

3 Estimation results

For the estimation of the model we use quarterly data of the Hungarian economy from the second quarter of 2001 to the last quarter of 2013. Eleven observed variables are selected – eight for the domestic economy and three for the foreign sector (represented by countries of the euro area). The variables are gross domestic product per economically active population, CPI inflation, nominal interest rate, real effective exchange rate, unemployment rate, average hours per worker, real wage, number of vacancies per economically active population and for the foreign economy: output, inflation and nominal interest rate. These data sets are acquired from the OECD and Bloomberg databases. All of them are seasonally adjusted and filtered. Domestic and foreign inflations and interest rates are demeaned and the remaining seven are detrended using Hodrick-Prescott filter with smoothing parameter $\lambda = 1600$. The Dynare 4.4.2 toolbox for MATLAB was used for the estimations.

We calibrate the model using some values from the available literature and some calculated from the data. The discount factor β is set to a frequently used value 0.99. The share of foreign goods in the domestic consumption is calculated as the import share on GDP and set to 0.4. The share of labour in production is set to 66%. Debt elasticity of financial risk premium is set to 0.001 according to Jakab and Kónya [4]. The steady state of the scaling factor, κ^v is set to 0.05 as in Lubik [6]. The scaling factor to disutility of work, κ^h is set in a way to get 1/3 of hours worked from the total amount of time. The steady state value of unemployment is calculated as the average value in the observed period and set to 0.082. The steady state of job finding rate ($\bar{f} = 0.18$) is taken from Jakab and Kónya [4]. Given the steady state equations, the separation rate is calculated as $\rho^x = \frac{\bar{f}(1-\bar{N})}{\bar{N}(1-\bar{f})} = 0.0196$.

Table 1 contains the prior densities of the estimated parameters. The prior for the deep-habit parameter is set like in Jakab and Kónya [4] to detect possible differences between the estimation results. The inverse of Frisch elasticity and the elasticity of substitution between domestis and foreign goods are set to 1. The bargaining power of firms is set to 0.8 with respect to the fact, that Hungarian trade unions are weak. With no information on the elasticity of matching, we decided to use the value in the middle of its (0,1) interval. As in Albertini et al. [1] or Lubik [6], the elasticity of vacancy adjustment is set to show linear hiring cost ($e = 1$). However, the large variance allows the cost function to change its form to concave or convex. We kept the price and wage setting parameters as they are stated in Albertini et al. [1] and combined the monetary policy parameters from Albertini et al. [1] and Jakab and Kónya [4].

Description	Symbol	Prior density	Posterior mean	90% HPDI
Habit	ϑ	$\beta(0.6, 0.15)$	0.4621	[0.3453;0.5794]
Inverse of Frisch elasticity	ϕ_h	$\Gamma(1, 0.2)$	1.2297	[1.0914;1.3550]
El. of substitution (dom. & for.)	η	$\Gamma(1, 0.2)$	0.5275	[0.4020;0.6613]
Bargaining power of firms	ξ	$\beta(0.8, 0.1)$	0.9252	[0.8515;0.9956]
Elasticity of matching	ν	$\beta(0.5, 0.2)$	0.7475	[0.6224;0.8646]
Elasticity of vacancy creation	e	$\Gamma(1, 0.5)$	3.4066	[2.7184;4.0442]
Price and wage setting				
Backward looking price (dom. good)	γ_H	$\beta(0.75, 0.1)$	0.7814	[0.6835;0.8719]
Backward looking price (for. good)	γ_F	$\beta(0.75, 0.1)$	0.8224	[0.7387;0.9329]
Backward looking wage parameter	γ_W	$\beta(0.75, 0.1)$	0.6582	[0.5389;0.7805]
Price adjustment cost (dom. good)	ψ_H	$\Gamma(50, 15)$	33.7960	[22.9016;46.8907]
Price adjustment cost (for. good)	ψ_F	$\Gamma(50, 15)$	71.4257	[51.8088;85.4372]
Wage adjustment cost	ψ_W	$\Gamma(50, 15)$	7.2486	[4.9227;9.5241
Monetary policy				
Interest rate smooth.	ρ_r	$\beta(0.5, 0.1)$	0.4645	[0.3064;0.5982]
Inflation	ρ_π	$\Gamma(1.5, 0.25)$	3.5043	[2.9219;4.0426]
Output gap	ρ_Y	$\mathcal{N}(0.25, 0.1)$	0.1359	[0.0490;0.2315]
Difference of output	$\rho_{\Delta Y}$	$\mathcal{N}(0.25, 0.1)$	0.2068	[0.0576;0.3313]
Exchange rate	ρ_e	$\mathcal{N}(0.25, 0.1)$	0.5216	[0.3545;0.6731]

Table 1 Estimation results (parameters)

		Prior density: $\beta(0.5, 0.2)$			Prior density: $\Gamma^{-1}(0.01, \infty)$	
Description	Persist.	Post. mean	90% HPDI	St. dev.	Post. mean	90% HPDI
Productivity	ρ_a	0.5686	[0.4327;0.7067]	σ_{ϵ_a}	0.0115	[0.0096;0.0133]
UIP	ρ_{uip}	0.7590	[0.6828;0.8401]	$\sigma_{\epsilon_{uip}}$	0.0066	[0.0047;0.0085]
Preference	ρ_c	0.4930	[0.3850;0.6076]	σ_{ϵ_c}	0.0756	[0.0592;0.0918]
Cost-push	ρ_H	0.3061	[0.1541;0.4429]	σ_{ϵ_H}	0.4854	[0.3316;0.6434]
Monetary	ρ_m	0.1193	[0.0335;0.1999]	σ_{ϵ_m}	0.0243	[0.0175;0.0309]
Matching	ρ_χ	0.4126	[0.2879;0.5372]	σ_{ϵ_χ}	0.1184	[0.0990;0.1375]
Bargaining	ρ_ξ	0.4373	[0.2321;0.6740]	σ_{ϵ_ξ}	0.0738	[0.0043;0.1499]
Vacancy	ρ_v	0.6508	[0.4653;0.8291]	σ_{ϵ_v}	0.0100	[0.0023;0.0178]
Disutility of work	ρ_h	0.1415	[0.0333;0.2450]	σ_{ϵ_h}	0.0999	[0.0718;0.1269]
For. output	ρ_{Y^*}	0.7539	[0.6477;0.8599]	$\sigma_{\epsilon_{Y^*}}$	0.0048	[0.0040;0.0055]
For. inflation	ρ_{π^*}	0.5928	[0.4606;0.7320]	$\sigma_{\epsilon_i^*}$	0.0017	[0.0014;0.0019]
For. int. rate	ρ_{i^*}	0.6889	[0.6188;0.7621]	$\sigma_{\epsilon_{\pi^*}}$	0.0041	[0.0034;0.0048]

Table 2 Estimation results (shocks)

Table 1 also contains the posterior means and 90% highest posterior density intervals (HDPI) of the estimated parameters. The estimated value of habit parameter (0.46) is much smaller than the value estimated by Jakab and Kónya [4], 0.68. One of the reasons for this difference could be the presence of the financial crisis period in our estimation. During recessions the households do not want to or cannot afford to consume as much as in the previous periods, therefore the weight of the past consumption declines. The bargaining power of firms is estimated to be really high (0.93), as was originally expected. This means, that firms get almost all surplus generated by the filled jobs. Also employees have practically no vote when negotiating wages. Elasticity of vacancy creation is estimated relatively high with value 3.4. This means, that the vacancy creation cost function is convex, thus each additional posted vacancy costs more, than the previous. The wage and price adjustment parameters show, that it is much easier (cheaper) to change the wage. Therefore the wages adapt quicker than prices. From the two prices, it is less costly to adjust the prices of the domestically produced goods. This is probably because the importers have to take into consideration the varying exchange rates when changing the prices. The interest rate smoothing parameter in the Taylor rule is relatively small, only 0.46 and the reaction to inflation is high, 3.5. The estimations of Jakab and Kónya [4] are 0.7 and 1.53 respectively. However, this significant difference can come from the different form of the Taylor rule, since their monetary rule does not include the gap and the first difference of the output. The estimation results of the persistence

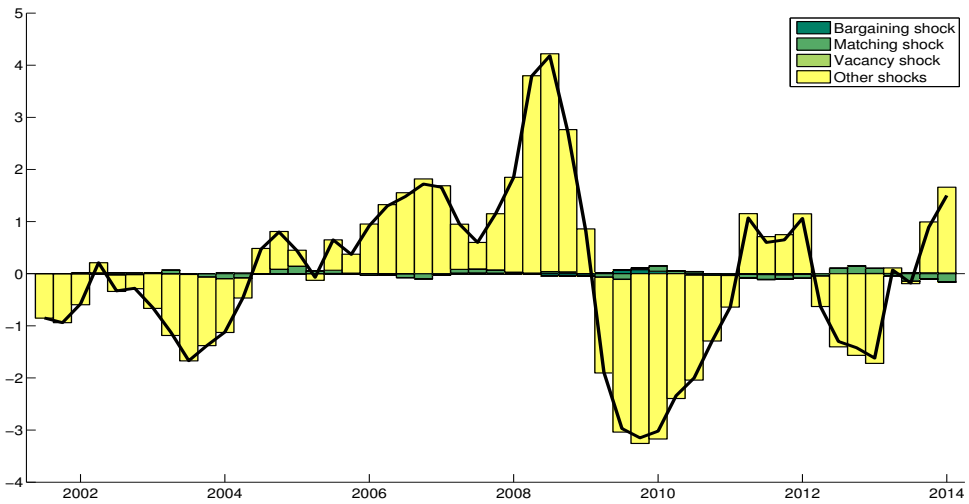


Figure 1 Variance decomposition of output

and deviation of shocks are presented in table 2. The monetary shock is the less persistent. It takes only half year for this shock to reach 10% of its initial value, while it takes more than five periods for the productivity shock to do the same. The cost-push shock is by far the most volatile. On the other hand, the volatilities of foreign shocks are the lowest.

4 Conclusion

The estimation of parameters seems to be reasonably consistent with the literature. We find rather high values of price frictions in the Hungarian economy, while the wage frictions are somewhat smaller. The model also captures the low participation rate in trade unions. Furthermore, we get similar results of the historical shock decomposition of output as Albertini et al. [1], presented in figure 1. The model shows almost no reaction of the output to the labour market shocks. Its fluctuation is given solely by the non-labour market shocks (in the figure depicted as the sum of the *other shocks*).

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0808/2013.

References

[1] Albertini, J., Kamber, G. and Kirker, M.: Estimated small open economy model with frictional unemployment. *Pacific Economic Review* **17**(2), 2012, 326–353.

[2] Ascari, G., Castelnuovo, E. and Rossi, L.: Calvo vs. Rotemberg in a trend inflation world: An empirical investigation. *Journal of Economic Dynamics and Control* **35**(11), 2011, 1852–1867.

[3] Calvo, G. A.: Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics* **12** (3), 1983, 383–398.

[4] Jakab, Z. and Kónya, I.: An open economy DSGE model with search-and-matching frictions: the case of Hungary. MNB Working Papers 10-2009, 2009.

[5] Lombardo, G. and Vestin, D.: Welfare implications of Calvo vs. Rotemberg pricing assumptions. *Economics Letters* **100**(2), 2008, 275–279.

[6] Lubik, T. A.: Estimating a search and matching model of the aggregate labor market. *Economic Quarterly* **95**(2), 2009, 101–120.

[7] Mortensen, D. T. and Pissarides, C. A.: Job creation and job destruction in the theory of unemployment. *Review of Economic Studies*, **61** (3), 1994, 397–415.

[8] Neugebauer, J. and Wesselbaum, D.: Staggered Wages, Sticky Prices, and Labor Market Dynamics in Matching Models. *Kiel Institute for the World Economy*, Kiel Working Paper No. 1608, 2010

[9] Němec, D.: Labour market frictions in a small open economy model of the Czech Republic. In: RAMK, J. and STAVREK, D. (eds.) *Proceedings of 30th International Conference Mathematical Methods in Economics*. Karviná: Silesian University, School of Business Administration, 2012, 634–639.

[10] Nisticó, S.: The welfare loss from unstable inflation. *Economics Letters* **96**(1), 2007, 51–57

[11] Rotemberg, J. J.: Cyclical wages in a search-and-bargaining model with large firms. In: *NBER International Seminar on Macroeconomics, 2006*, University of Chicago Press, 2008, 65–114.

[12] Rotemberg, J. J.: Monopolistic price adjustment and aggregate output. *The Review of Economic Studies*, **49** (4), 1982, 517–531.

[13] Zanetti, F.: A non-Walrasian labor market in a monetary model of the business cycle. *Journal of Economic Dynamics and Control*, **31**, 2007, 2413–2437.

A dynamic analysis of causality between prices of corn, crude oil and ethanol

Monika Papież[1]

Abstract. The objective of the paper is to analyse causality among prices of corn, crude oil and ethanol. The analysis conducted in this paper is a dynamic one, and the data used consist of weekly futures prices of crude oil, corn, and ethanol from January 5, 2007 till April 11, 2014. The assessment of causal links between prices of corn, crude oil and ethanol is carried out with the use of rolling regression applied to augmented-VAR framework proposed by Toda and Yamamoto [22]. The application of the rolling regression procedures into the modified Wald (MWALD) causality test allows for the investigation of the persistence of stability in causal relations between analysed prices. The results obtained indicate that the linkages between energy prices and agricultural commodity prices change in the period analysed. The results of Granger causality tests reveal that in the analysed period the price of corn influences the price of energy (crude oil and ethanol). Also crude oil prices influence corn prices and ethanol prices. However, the influence of ethanol prices on crude oil prices and corn prices has not been observed.

Keywords: Granger causality, rolling regression, Toda -Yamamoto tests, commodity prices.

JEL Classification: C32, Q13, Q41
AMS Classification: 91B84, 62M10

1 Introduction

Limited fossil fuel resources and the fact that the demand for them is growing continuously coupled with the economic development and excessive pollution of the environment lead to the increase in the significance of renewable energy source. Additionally, the European Commission has set an overall binding target to satisfy 20% of the EU energy needs from renewable sources, such as biomass, hydro, wind and solar power by 2020. As part of the overall target, each member state has to achieve at least 10% of their transport fuel consumption from renewable sources (including biofuels). The factors mentioned above have resulted in the growth in the production of ethanol in the period 2007 - 2013 by about 70%. However, such increased demand for ethanol fuel translates into greater demand for corn, which, in turn, increases the prices of corn. On the other hand, the increase in the production of biofuels changes the structure of energy sources, which affects the prices of fossil fuel, including crude oil prices. What is more, Kilian and Park [8] claim that the price of crude oil has the greatest influence on food prices because its increasing price raises both transport costs and food production costs through the growth of fuel costs for mechanized farming. Additionally, growing prices of crude oil increase the economic motivation for the production of biofuels (corn, soybean, sugar cane, oil palm, etc.).

Hence, one of the most important effects of the growing biofuel production has been the change in the nature of the linkages between agricultural commodity markets and energy markets. Thus, it is interesting to investigate how the prices of biofuels affect fossil fuel prices and food prices and vice versa. In this study we analyse causality between corn prices representing the food prices, crude oil prices representing the fossil fuel prices, and ethanol prices representing the biofuel prices.

The investigation of related issues can be seen in numerous recent studies, although their conclusions are inconclusive. Some researchers analyse only the relations between food and fossil fuel prices generally ignoring biofuel prices. Some studies confirm the linkages between food prices and crude oil prices (e.g: Chen et al. [1], Ciaian and Kancs [2], Ciaian and Kancs [3], Harii et al. [7], Natanelov et al. [11], Nazlioglu and Soytas [14]; Nazlioglu [13], and Papież and Śmiech [16]). Other empirical studies report no evidence regarding the oil–food price nexus, thereby supporting the neutrality hypothesis. Nazlioglu and Soytas [15] and Zhang et al. [27] find agricultural commodity prices to be neutral to the effects of oil price changes in the long run.

Subject literature contains a large number of studies on the linkages between the prices of energy sources (fossil fuels and biofuels) and the prices of food. Many of these studies use time-series econometric techniques to quantify the relations between oil, ethanol, and food prices in levels (e.g., Kristoufek et al. [9], McPhail [10],

[1] Cracow University of Economics, Rakowicka 27, 31-510 Kraków, Poland e-mail: papiezm@uek.krakow.pl.

Natanelov et al. [12], Qiu et al. [17], Saghaian [19], Serra et al. [20], Wixson and Katchova [24], Zhang et al. [26], and Zhang et al. [27]) or their volatility interactions (e.g., Gardebroek and Hernandez [5], Haixia and Shiping [6], Trujillo-Barrera et al. [23]). However, to the best of our knowledge, dynamic causality in the crude oil–corn–ethanol–nexus analysed with the use of the rolling regression procedures applied into the modified Wald (MWALD) causality test has not been addressed in any of them yet.

The objective of this study is to investigate dependencies among prices of corn, crude oil and ethanol, using weekly futures data spanning from January 5, 2007 to April 11, 2014. The analysis of dependencies has a dynamic nature and focuses on Granger causality between the variables. The assessment of causal links between the variables is carried out with the use of rolling regression applied to augmented-VAR framework proposed by Toda and Yamamoto [22]. The application of the rolling regression procedures into the modified Wald (MWALD) causality test allows for the investigation of the persistence of stability in causal relations between analysed prices.

The analysis conduced allows us to verify the following hypotheses:

- The dependencies between the prices of energy sources and the food prices are not stable in time.
- The prices of biofuels or fossil fuels affect food prices in the short run.
- Food prices affect the prices of biofuels or fossil fuels in the short run.
- The prices of biofuels affect the prices of fossil fuels in the short run.

This paper contributes to the existing literature mostly due to the application of the dynamic analysis, which allows us to assess the stability of the dependencies between the variables. Additionally, incorporating the rolling regression procedure into causality tests provides more information on the issue of the crude oil–corn–ethanol–nexus. What is more, using a rolled window in the analysis makes it possible to indicate breaking points and facilitates their further interpretation.

The paper is organized as follows. Section 2 presents methodology. Data, the discussion of the methods used and the results obtained are given in Section 3, while Section 4 contains the main conclusions.

2 Methodology

The assessment of causal dependencies between prices of corn, crude oil and ethanol is carried out with the use of rolling regression applied to augmented-VAR framework proposed by Toda and Yamamoto [22] and developed by Rambaldi and Doran [18] and Zapata and Rambaldi [25]. This procedure avoids the problems of testing for Granger causality with respect to the power and size properties of unit root and co-integration tests (Zapata and Rambaldi [25]). The approach suggested by Toda and Yamamoto [22] applies the modified Wald (MWALD) causality test to the model $VAR(k+d_{\max})$, where k is the lag length of the system determined by information criteria (Akaike Information Criterion (AIC) or BIC), and $d_{\max}$ is the maximal order of integration. In a nutshell, the T–Y methodology involves the following stages. Firstly, the lag length (k) of the system VAR and the maximal order of integration $(d_{\max})$ are established. Secondly, the augmented $VAR(k+d_{\max})$s in levels are estimated. Next, for the model $VAR(k+d_{\max})$ the Wald test to the first k VAR coefficient matrix is performed to test for Granger causality. For testing the null hypothesis, Toda and Yamamoto [22] confirm that the Wald statistic has the asymptotic χ^2 - distribution with k degrees of freedom, regardless of whether the generating process is stationary (possibly around a linear trend) or cointegrated.

In our case, T–Y version of $VAR(k+d_{\max})$ can be written as:

$$\begin{aligned} OIL_t &= \alpha_{01} + \sum_{i=1}^{k} \alpha_{1i} OIL_{t-i} + \sum_{j=k+1}^{k+d_{\max}} \alpha_{2j} OIL_{t-j} + \sum_{i=1}^{k} \beta_{1i} CORN_{t-i} + \sum_{j=k+1}^{k+d_{\max}} \beta_{2j} CORN_{t-j} + \\ &+ \sum_{i=1}^{k} \gamma_{1i} ETHANOL_{t-i} + \sum_{j=k+1}^{k+d_{\max}} \gamma_{2j} ETHANOL_{t-j} + \varepsilon_{1t} \end{aligned} \tag{1}$$

$$CORN_t = \alpha_{02} + \sum_{i=1}^{k} \alpha_{3i} OIL_{t-i} + \sum_{j=k+1}^{k+d_{\max}} \alpha_{3j} OIL_{t-j} + \sum_{i=1}^{k} \beta_{3i} CORN_{t-i} + \sum_{j=k+1}^{k+d_{\max}} \beta_{4j} CORN_{t-j} + \\ + \sum_{i=1}^{k} \gamma_{3i} ETHANOL_{t-i} + \sum_{j=k+1}^{k+d_{\max}} \gamma_{4j} ETHANOL_{t-j} + \varepsilon_{2t} \tag{2}$$

$$ETHANOL_t = \alpha_{03} + \sum_{i=1}^{k} \alpha_{5i} OIL_{t-i} + \sum_{j=k+1}^{k+d_{\max}} \alpha_{6j} OIL_{t-j} + \sum_{i=1}^{k} \beta_{5i} CORN_{t-i} + \sum_{j=k+1}^{k+d_{\max}} \beta_{6j} CORN_{t-j} + \\ + \sum_{i=1}^{k} \gamma_{5i} ETHANOL_{t-i} + \sum_{j=k+1}^{k+d_{\max}} \gamma_{6j} ETHANOL_{t-j} + \varepsilon_{3t} \tag{3}$$

The directions of Granger causality can be detected by applying standard Wald tests to the first k VAR coefficient matrix. For example, for Eq. (1): $H_0 : \beta_{11} = \beta_{12} = ... = \beta_{1k} = 0$, implies that corn prices (CORN) do not Granger cause crude oil prices (OIL), and $H_0 : \gamma_{11} = \gamma_{12} = ... = \gamma_{1k} = 0$, implies that ethanol prices (ETHANOL) do not Granger cause crude oil prices (OIL).

The changes in dependencies between the variables over time are investigated with the use of rolling analysis (Smiech and Papież [21]). We apply the fixed window rolling regression to the level VAR model. The first model is built using the data covering observations from 1 to n, the second model covers observations from 2 to n+1, etc.. Every time AIC is used to fix the number of lags of VAR model. Next, we estimate the parameters of VAR models, and, finally, we use the MWALD test statistic to test Granger causality. This allows us to observe whether and how the dependencies between the variables change for consecutive rolling windows.

3 Data and empirical results

The data used in this study consist of weekly prices of crude oil (OIL), corn (CORN), and ethanol (ETHANOL) from the period between 5 January 2007 and 11 April 2014 (380 observations). The data used in the analysis include the prices of futures contracts traded on the New York Mercantile Exchange (NYMEX) and the Chicago Board of Trade (CBOT). The present study uses nominal data because weekly consumer price index is unavailable. The detailed description of variables and descriptive statistics for weekly time series data are presented in Table 1. Next, for the purpose of the study, all the variables are converted to their natural logarithm form.

Variable	Symbol	Unit	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis
Crude oil	NYMEX:CL	$/bbl	86.47	89.30	141.73	37.93	19.01	-0.21	3.35
Corn	CBOT:C	$/bu	5.20	4.87	8.14	3.09	1.43	0.31	1.69
Ethanol	CBOT:EH	$/gal	2.10	2.14	2.94	1.42	0.39	-0.09	2.09

Table 1 Summary statistics for weekly time series

To investigate the stationarity issue and the possible presence of unit roots in series, univariate analysis of each of the time series is carried out. Augmented Dickey–Fuller (ADF) unit root tests (Dickey and Fuller [4]) for individual time series and their differences are used. ADF statistic testing for unit root shows the rejection at 1% significance level, implying the stationarity of each return series for each analysed period[2]. The number of lags in the test is established using AIC criterion.

Conducting the analysis within the rolling regression requires obtaining the window size (VAR models with fixed sample size each time, i.e. a fixed window size). The VAR models are calculated for a rolling 104 observations (approximately 2 calendar years) time window by adding one observation to the end and removing the first observation and so on. That is, starting with observations 1–104, we calculate the first VAR model. Then, we calculate the VAR model for observations 2–105, 3–106, etc. Using AIC, we determine k – the number of lags in VAR models for each window.

[2] The results can be obtained from the author on request.

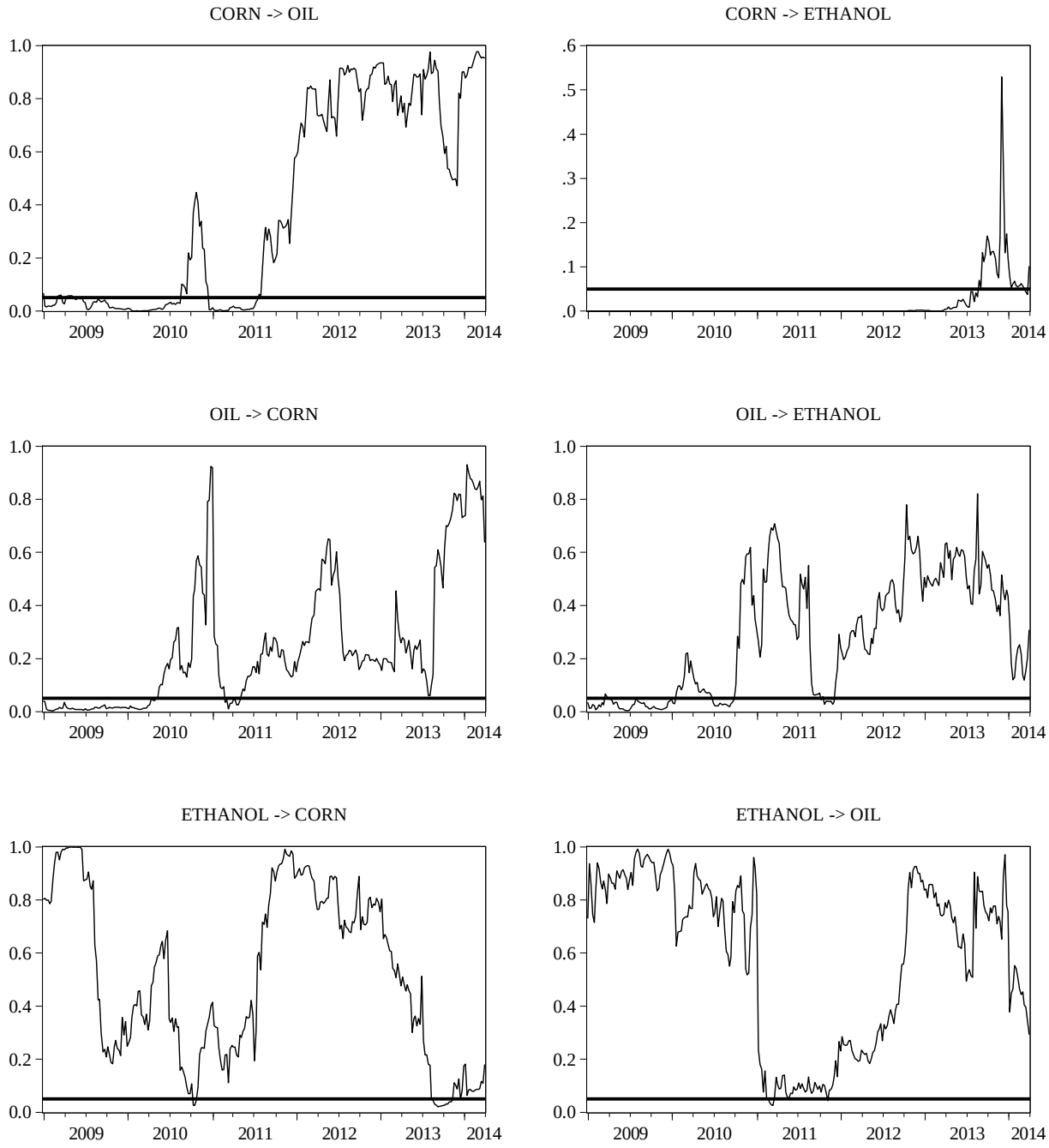


Figure 1 Rolling Granger causality test - the value of p-value of MWALD test.

Granger causality tests can be applied via a MWALD test statistic on the first *k* coefficients. Figure 1 presents *p*-value for Granger causality tests. The horizontal axis indicates the ending point of the window of analysis. (We report the test statistics on the last day of the rolling sample period from which they are derived) The first value represents the *p*-value for Granger causality tests for the model estimated for the period from 5 January 2007 to 26 December 2008. The last one represents the *p*-value for Granger causality tests in VAR estimated for the window 20 April 2012 – 11 April 2014. The horizontal line in the chart indicates the significance level of 5%. The values below this line mean that for a given subperiod variable A Granger causes variable B (A → B).

The results presented in Fig. 1 indicate that in the period analysed the variables influencing other variables change. The analysis of the results presented in Fig. 1 indicates that crude oil prices influence corn prices in the subperiods which start at the beginning of the analysis, that is January 2007 (that is the subperiod from January 2007 till December 2008), until the subperiod beginning in April 2008 (that is, the last dependence subperiod lasted from April 2008 till March 2010). It is the period of considerable increases in oil prices and their rapid drops connected with the global financial crisis after the collapse of Lehman Brothers. In the subperiods which

begin after April 2008, the influence of crude oil prices on corn prices is not observed. Similarly, corn prices influence crude oil prices in the subperiods beginning in January 2007. Their influence is longer, however, and the last subperiod for which past values of corn prices improve the forecasts of the crude oil prices is observed from September 2009 till August 2011. The results of the analysis indicate mutual dependence between crude oil prices and corn prices from the beginning of the analysis up to the first quarter of 2010. Later corn prices influence crude oil prices. No dependencies between agricultural commodity prices represented by corn prices and crude oil prices are found for the subperiods which begin in the second half of 2011.

Similarly, the results of the analysis presented in Fig. 1. indicate that crude oil prices influence ethanol prices for the subperiods beginning in January 2007 up to the subperiods beginning in October 2008 (that is, the last dependence period lasted from October 2008 till September 2010). However, the impact of ethanol prices on crude oil prices is not observed in the whole analysed period, which means that, within the energy market, Granger causality tests show that changes in the price of oil are an indicator of future changes in the price of ethanol. This relationship is unidirectional with changes in the price of ethanol unable to help predict future changes in the price of oil.

It can also be noticed that past values of corn prices improve the forecasts of ethanol prices from the beginning of the analysis up to the subperiods which begin in the third quarter of 2011 and last up to the third quarter of 2013. Corn prices do not influence ethanol prices in the subperiods which end from the fourth quarter of 2013 on, whereas in the whole period analysed significant causal relationships between the ethanol prices and corn prices are not observed.

4 Conclusion

The objective of the study is a dynamic assessment of dependencies between prices of corn, crude oil and ethanol using weekly data spanning from January 2007 to April 2014. The analysis which uses the rolling regression to augmented VAR models allows us to disprove or approve the hypotheses posed at the beginning. The results confirm the hypothesis stating that the dependencies between the prices of energy sources and the food prices change in time.

Because the results obtained indicate that dependencies are not stable in time, it is not possible to approve or disprove the next two hypotheses. The results of our analysis indicate that food prices represented by the corn prices influence the prices of energy sources. The corn prices affect fossil fuel prices (that is, crude oil prices) only up to the middle of 2010, while they affect biofuel prices (that is, ethanol prices) up to the third quarter of 2013. In the later period the impact of corn prices on energy sources prices is not observed. Similarly, using monthly data from 1995:01 to 2010:12, Wixson and Katchova [24] show that changes in the prices of corn can be a leading indicator of changes in the prices of oil and ethanol.

The results of Granger causality tests indicate that changes in crude oil prices can be a leading indicator of changes in corn prices only up to the first quarter of 2010 and in ethanol prices up to the third quarter of 2010. In the later period the impact of crude oil prices on corn prices and ethanol prices is not observed.

The results obtain do not confirm the hypothesis stating that the prices of biofuels affect the prices of fossil fuels in the short run. The study indicates that the price of biofuels represented by ethanol prices does not influence either fossil fuel prices represented by crude oil prices or food prices represented by corn prices.

Additionally, it can be concluded that, from the third quarter of 2010 on, the results indicate that there are no causal relations between fossil fuel prices (represented by crude oil prices) and biofuel prices (represented by ethanol prices). Zhang et al. [26] find a similar lack of relationship between ethanol prices and crude oil prices in the period of the ethanol boom (2000-2007), although McPhail [10], who uses monthly data from the period of 1994:01–2010:02, shows that real ethanol prices Granger cause real oil prices and vice versa.

Acknowledgements

Supported by the grant No. 2012/07/B/HS4/00700 of the Polish National Science Centre.

References

[1] Chen, S., Kuo, H., and Chen, C.: Modeling the relationship between the oil price and global food prices. *Applied Energy* **87** (2010), 2517–2525.

[2] Ciaian, P., and Kancs, A.: Food, energy and environment: is bioenergy the missing link? *Food Policy* **36** (2011a), 571–580.

[3] Ciaian, P., Kancs, A.,: Interdependencies in the energy–bioenergy–food price systems: a cointegration analysis. *Resource and Energy Economics* **33** (2011b), 326–348.

[4] Dickey, D., and Fuller, W.: Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74** (1979), 427–431.
[5] Gardebroek, C., and Hernandez, M. A.: Do energy prices stimulate food price volatility? Examining volatility transmission between US oil, ethanol and corn markets. *Energy Economics* **40** (2013), 119-129.
[6] Haixia, W., and Shiping, L.: Volatility spillovers in China's crude oil, corn and fuel ethanol markets. *Energy Policy* **62** (2013), 878-886.
[7] Harri, A., Nalley, L., and Hudson, D.: The relationship between oil, exchange rates, and commodity prices. *Journal of Agricultural and Applied Economics* **41** (2009), 501–510.
[8] Kilian, L., and Park, C.: The impact of oil price shocks on the U.S. stock market. *International Economic Review* **50** (2009), 1267-1287.
[9] Kristoufek, L., Janda, K., and Zilberman, D.: Mutual Responsiveness of Biofuels, Fuels and Food Prices. *Working paper*, Centre for Applied Macroeconomic Analysis. Australian National University, Canberra, 2012.
[10] McPhail, L. L.: Assessing the impact of US ethanol on fossil fuel markets: A structural VAR approach. *Energy Economics* **33**(6) (2011), 1177-1185.
[11] Natalenov, V., Alam, M.J., McKenzie, A.M., and Van Huylenbroeck, G.: Is there co-movement of agricultural commodities futures prices and crude oil? *Energy Policy* **39** (2011), 4971–4984.
[12] Natanelov, V., McKenzie, A. M., and Van Huylenbroeck, G.: Crude oil–corn–ethanol–nexus: A contextual approach. *Energy Policy* **63** (2013), 504-513.
[13] Nazlioglu, S.: World oil and agricultural commodity prices: evidence from nonlinear causality. *Energy Policy* **39** (2011), 2935–2943.
[14] Nazlioglu, S., and Soytas, U.: World oil prices and agricultural commodity prices: evidence from an emerging market. *Energy Economics* **33** (2011), 488–496.
[15] Nazlioglu, S., and Soytas, U.: Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. *Energy Economics* **34** (2012), 1098–1104.
[16] Papież, M., and Śmiech, S.: Causality in mean and variance between returns of crude oil and metal prices, agricultural prices and financial market prices. In: *Proceedings of 30th International Conference Mathematical Methods in Economics* (Ramík, J. and Stavárek, D. (eds.)). Karviná: Silesian University, School of Business Administration, 2012, 675-680.
[17] Qiu, C., Colson, G., Escalante, C., and Wetzstein, M.: Considering macroeconomic indicators in the food before fuel nexus. *Energy Economics* **34** (2012), 2021–2028.
[18] Rambaldi, A.N., and Doran H.E.: Testing for Granger non-causality in cointegrated system made easy. *Working papers in Econometrics and Applied Statistics*, **88**, Department of Econometrics, University of New England, 1996.
[19] Saghaian, S.H.: The impact of the oil sector on commodity prices: correlation or causation? *Journal of Agricultural and Applied Economics* **42**, (2010), 477–485.
[20] Serra, T., Zilberman, D., Gil, J. M., and Goodwin, B. K.: Nonlinearities in the US corn‐ethanol‐oil‐gasoline price system. *Agricultural Economics* **42**(1) (2011), 35-45.
[21] Śmiech, S., and Papież, M.: Fossil fuel prices, exchange rate, stock market: a dynamic causality analysis on the European market. *Economics Letters* **118** (2013), 199-202.
[22] Toda, H. Y., and Yamamoto, T.: Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* **66** (1995), 225-250.
[23] Trujillo-Barrera, A., Mallory, M., and Garcia, P.: Volatility Spillovers in US Crude Oil, Ethanol, and Corn Futures Markets. *Journal of Agricultural and Resource Economics* **37**(2) (2012), 247–262.
[24] Wixson, S.E., and Katchova, A.E.: Price asymmetric relationships in commodity and energy markets. In: *Paper presented at the 123rd European Association of Agricultural Economists' Seminar, Price Volatility and Farm Income Stabilisation*, Dublin, February 23–24, 2012.
[25] Zapata, H.O., and Rambaldi, A.N.: Monte Carlo evidence on cointegration and causation. *Oxford Bulletin of Economics and Statistics* **59**(2) (1997), 285-298.
[26] Zhang, Z., Lohr, L., Escalante, C., and Wetzstein, M.: Ethanol, corn, and soybean price relations in a volatile vehicle-fuels market. *Energies* **2**(2) (2009), 320-339.
[27] Zhang, Z., Lohr, L., Escalante, C., and Wetzstein, M.: Food versus fuel: what do prices tell us? *Energy Policy* **38** (2010), 445–451.

Probability of fuzzy events

Ondřej Pavlačka[1], Pavla Rotterová[2]

Abstract. In economic practice, we often deal with events that are defined only vaguely. Such indeterminate events can be modelled by fuzzy sets. In the paper, we examine two main ways of expressing a probability of fuzzy events that are proposed in the literature. We study their mathematical properties and discuss their interpretation. We conclude that none of the approaches gives an appropriate probability of a fuzzy event, and thus, that the question "How should the probability of a fuzzy event be expressed?" is still an open problem.

Keywords: fuzzy probability spaces, fuzzy events, probability measure, decision making under risk.

JEL classification: C44
AMS classification: 90B50

1 Introduction

In the models of decision making under risk, a probability space is considered, i.e. we are able to assign probabilities to some precisely defined random events, like "an interest rate is less than 1.5 % p.a.", "a loss is greater than 100 000 CZK", etc. However, in economic practice we often deal with events that are defined only vaguely, like "a low interest rate", "an appropriate revenue", "a big loss", etc. Such indeterminate events can be adequately modelled by fuzzy sets on the universal set (see [7]). As we often need to estimate the probability of such events, we need to extend the given probability space to the case of fuzzy events.

Extending the given probability space to the case of fuzzy events means: first, to determine which fuzzy sets on the corresponding universal set are eligible to be fuzzy events, and consequently, to define the way of expressing probabilities of such fuzzy events.

In the literature, fuzzy events are typically defined as the fuzzy sets whose α-cuts are random events. Two main approaches to expressing their probabilities were established. The common way was introduced by Zadeh [8] in 1968. He defined the probability of a fuzzy event as the expected value of its membership function, i.e. the probability of a fuzzy event is a real number from the unit interval. Another way, proposed by Yager [5] in 1979 and independently by Talašová and Pavlačka [4] in 2006, consists in expressing the probability of a fuzzy event by a *fuzzy probability* - a fuzzy set defined on [0,1] whose membership function is derived from the probabilities of the α-cuts of the fuzzy event. The aim of the paper is to examine if both the approaches are appropriate to be used in practice.

The paper is organized as follows. In Section 2, a definition of a probability space and some important properties of a probability measure are recalled. Section 3 is devoted to basic notions from fuzzy sets theory. In the next two sections, we examine the two different approaches to expressing the probability of fuzzy events. Finally, some concluding remarks are given in Section 6.

2 A probability space and the properties of a probability measure

A probability space, introduced by Kolmogorov [2] in 1933, is an ordered triple $(\Omega, \mathcal{A}, p)$, where Ω denotes a non-empty set of all elementary events (future states of the world), $\mathcal{A}$ represents the set of all considered random events ($\mathcal{A}$ forms a σ-algebra of subsets of Ω), and $p : \mathcal{A} \to [0, 1]$ is a probability measure that assigns to each random event $A \in \mathcal{A}$ its probability $p(A) \in [0, 1]$ satisfying the following conditions:

[1]Palacký University Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, ondrej.pavlacka@upol.cz
[2]Palacký University Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, pavla.melicherikova01@upol.cz

1.
$$p(\Omega) = 1, \tag{1}$$

2. for any $A_1, A_2, \ldots \in \mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for any $i, j \in \mathbb{N}$, $i \neq j$:
$$p\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} p(A_n). \tag{2}$$

Any probability measure p possesses also the following well known properties that play a significant role in practical applications:

1. $p(\emptyset) = 0$,
2. for all $A, B \in \mathcal{A}$, $A \subseteq B$: $p(A) \leq p(B)$,
3. for all $A, B \in \mathcal{A}$: $p(A \cup B) = p(A) + p(B) - p(A \cap B)$,
4. for any $A \in \mathcal{A}$: $p(A^c) = 1 - p(A)$,
5. for any $A_1, \ldots, A_n \in \mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, \ldots, n\}$, $i \neq j$, and $\bigcup_{i=1}^{n} A_i = \Omega$: $\sum_{i=1}^{n} p(A_i) = 1$.

Later in the paper, we will study retaining of these properties also for the probabilities of fuzzy events.

3 Fuzzy sets

In this section, let us briefly introduce basic notions of fuzzy sets theory. We will recall the definitions of a fuzzy set and its characteristics, inclusion between fuzzy sets, and basic operations with fuzzy sets.

A fuzzy set A on a nonempty set Ω is characterized by its membership function $\mu_A : \Omega \to [0, 1]$. The family of all fuzzy sets on Ω will be denoted by $\mathcal{F}(\Omega)$. By $Core\, A$ and $Supp\, A$, we denote a core of A, i.e. $Core\, A := \{\omega \in \Omega \mid \mu_A(\omega) = 1\}$, and a support of A, i.e. $Supp\, A := \{\omega \in \Omega \mid \mu_A(\omega) > 0\}$, respectively. For any $\alpha \in (0, 1]$, A_α means an α-cut of A, i.e. $A_\alpha := \{\omega \in \Omega \mid \mu_A(\omega) \geq \alpha\}$.

Let us note that any crisp set $A \subseteq \Omega$ can be viewed as a fuzzy set of a special kind; the membership function μ_A coincides in such a case with the characteristic function χ_A of A. For a crisp set A, $Supp\, A = A$, and $A_\alpha = A$ for all $\alpha \in (0, 1]$.

A fuzzy set $A \in \mathcal{F}(\Omega)$ is said to be a subset of $B \in \mathcal{F}(\Omega)$, we will denote it by $A \subseteq B$, if $\mu_A(\omega) \leq \mu_B(\omega)$ holds for all $\omega \in \Omega$. Obviously, $A \subseteq B$ if and only if $A_\alpha \subseteq B_\alpha$ for any $\alpha \in (0, 1]$.

The intersection and union of two fuzzy sets $A, B \in \mathcal{F}(\Omega)$ are defined as fuzzy sets $A \cap B, A \cup B \in \mathcal{F}(\Omega)$ whose membership functions are for all $\omega \in \Omega$ given by $\mu_{A \cap B}(\omega) = \min\{\mu_A(\omega), \mu_B(\omega)\}$ and $\mu_{A \cup B}(\omega) = \max\{\mu_A(\omega), \mu_B(\omega)\}$, respectively. Since the minimum and maximum operations are used, $(A \cap B)_\alpha = A_\alpha \cap B_\alpha$ and $(A \cup B)_\alpha = A_\alpha \cup B_\alpha$ hold for all $\alpha \in (0, 1]$.

A complement of a fuzzy set A is a fuzzy set A^c whose membership function is for all $\omega \in \Omega$ given by $\mu_{A^c}(\omega) = 1 - \mu_A(\omega)$. Obviously, $(A^c)_\alpha = (A_\alpha)^c$ for any $\alpha \in (0, 1]$.

4 Extension of a given probability space to the case of fuzzy events

Now, let us assume that a probability space $(\Omega, \mathcal{A}, p)$ is given and needs to be extended to the case of fuzzy events defined on Ω. In this section, we will analyse the most common way of such extension that was proposed by Zadeh [8] (in fact, Zadeh [8] considered $\Omega = \mathbb{R}^n$ and $\mathcal{A} = \mathcal{B}_n$, where $\mathcal{B}_n$ denotes the σ-algebra of Borel sets in $\mathbb{R}^n$, but the proposed extension can be without any modifications applied to the case of general Ω and general $\mathcal{A}$).

As it was mentioned in Introduction, first of all, we have to determine which fuzzy sets on Ω are eligible to be fuzzy events. According to Zadeh [8], *a fuzzy event* is a fuzzy set $A \in \mathcal{F}(\Omega)$ whose membership function is $\mathcal{A}$-measurable, i.e. $A_\alpha \in \mathcal{A}$ for any $\alpha \in (0, 1]$. The family of all such fuzzy events, let us denote it by $\mathcal{A}_F$, forms a σ-algebra of fuzzy sets on Ω (see Negoita and Ralescu [3]), i.e.

1. $\Omega \in \mathcal{A}_F$,
2. for any $A_1, A_2 \ldots \in \mathcal{A}_F$: $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_F$,
3. for any $A \in \mathcal{A}_F$: $A^c \in \mathcal{A}_F$,

Note that $\mathcal{A} \subseteq \mathcal{A}_F$.

Now, the probability measure p needs to be extended to the case of fuzzy event. Let us denote the extension by p_F. Zadeh [8] defined the probability $p_F(A)$ of a fuzzy event $A \in \mathcal{A}_F$ as the expected value of its membership function μ_A, i.e. by the Lebesgue-Stieltjes integral

$$p_F(A) := E(\mu_A) = \int_{\Omega} \mu_A(\omega) dp. \tag{3}$$

The existence of the above Lebesgue-Stieltjes integral follows directly from the assumption that μ_A is $\mathcal{A}$-measurable. It can be easily seen that for any crisp random event $A \in \mathcal{A}$, $p_F(A) = p(A)$. Talašová and Pavlačka [4] showed that the probability of a fuzzy event A given by (3) can be equivalently expressed as follows:

$$p_F(A) = \int_0^1 p(A_\alpha) d\alpha.$$

It is obvious from (3) that $p_F : \mathcal{A}_F \to [0, 1]$. Furthermore, Negoita and Ralescu [3] showed that p_F fulfills also the two conditions for a probability measure, i.e. $p_F(\Omega) = 1$, and for any $A_1, A_2, \ldots \in \mathcal{A}_F$ such that $A_i \cap A_j = \emptyset$ for any $i, j \in \mathbb{N}$, $i \neq j$: $p_F\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} p_F(A_n)$. As also $\mathcal{A}_F$ forms a σ-algebra of fuzzy sets on Ω, they called the ordered triple $(\Omega, \mathcal{A}_F, p_F)$ *a fuzzy probability space.*

Let us show now that p_F possesses also the other properties of the probability measure mentioned in Section 2. Obviously, $p_F(\emptyset) = p(\emptyset) = 0$. Zadeh [8] showed that if $A, B \in \mathcal{A}_F$, $A \subseteq B$, then $p_F(A) \leq p_F(B)$, and that for all $A, B \in \mathcal{A}_F$, $p_F(A \cup B) = p_F(A) + p_F(B) - p_F(A \cap B)$. Furthermore, for any $A \in \mathcal{A}_F$:

$$p_F(A^c) = E(\mu_{A^c}) = E(1 - \mu_A) = 1 - E(\mu_A) = 1 - p_F(A).$$

The fifth property can be generalized as follows:

Proposition 1. *If $A_1, \ldots, A_n \in \mathcal{A}_F$ form* a fuzzy partition of Ω, *i.e.* $\sum_{i=1}^{n} \mu_{A_i}(\omega) = 1$ *for all* $\omega \in \Omega$, *then*

$$\sum_{i=1}^{n} p_F(A_i) = 1. \tag{4}$$

Proof.

$$\sum_{i=1}^{n} p_F(A_i) = \sum_{i=1}^{n} E(\mu_{A_i}) = E\left(\sum_{i=1}^{n} \mu_{A_i}\right) = E(\mathbf{1}) = 1,$$

where $\mathbf{1}$ is a random variable on Ω such that $\mathbf{1}(\omega) = 1$ for all $\omega \in \Omega$. □

Thus, we can see that from a mathematical point of view, the mapping p_F given by (3) represents a correct extension of the probability measure p. However, the question is if for a fuzzy event $A \in \mathcal{A}_F$ which is not crisp, the value $p_F(A)$ represents its probability in a common sense. Let us discuss now the problem in more detail.

The probability $p(A)$ of a uniquely determined crisp event A is commonly interpreted as a measure of the chance that the event A occurs in the future. For instance, if A denotes an event "the interest rate will be less or equal to 2 % p.a." and $p(A) = 0.5$, then we know that there is the exactly same chance that the interest rate will be less or equal to 2 % p.a. or that it will be greater than 2 % p.a. In other words, if we have the possibility to infinitely many times repeat the process, we can expect that the event A will occur in 50 % cases.

However, if we have a vaguely defined event "the interest rate will be low", expressed by a fuzzy set A, and if $p_F(A) = 0.5$, the only thing we know is that we can expect in the future the interest rate i % p.a. such that $\mu_A(i) = 0.5$, i.e. the expected possibility that i is "small" is equal to 0.5. And this is the completely different meaning than the common interpretation of the probability of a crisp event; it is

not related to the chance that A will occur in the future. In fact, if an interest rate i^* % p.a. such that $\mu_A(i^*) \in (0,1)$ will occur in the future, we will not be able even to uniquely decide whether A occurred or not, i.e. whether the interest rate i is or is not "small". Let us note that the only way the value $p_F(A)$ would have the same meaning as the probability of a crisp event is that the membership degree $\mu_A(u)$ is interpreted as the probability that an object u belongs to A. Such interpretation of the membership degrees was proposed by Hisdal [1].

Another problem is that the value $p_F(A)$ says nothing about the fuzziness of a fuzzy event $A \in \mathcal{A}_F$. This is the reason why Yager [5], and Talašová and Pavlačka [4] proposed the idea that a probability of a fuzzy event should be also fuzzy. This approach will be analysed in the next section.

Example 1. Let us consider the following example (its idea is taken from Zadeh [6]): "An urn contains 20 balls $b_1, b_2, \ldots, b_{20}$ of various sizes. What is the probability that a ball drawn at random is large?"

A discrete fuzzy set of large balls B_L is given by the following formula:

$$B_L = \left\{ {}^{0.5}|_{b_1}, {}^{0}|_{b_2}, {}^{1}|_{b_3}, {}^{0.2}|_{b_4}, {}^{1}|_{b_5}, {}^{0.7}|_{b_6}, {}^{0.3}|_{b_7}, {}^{0}|_{b_8}, {}^{1}|_{b_9}, {}^{0.4}|_{b_{10}}, {}^{1}|_{b_{11}}, \right. \\ \left. {}^{0.6}|_{b_{12}}, {}^{0.1}|_{b_{13}}, {}^{0.8}|_{b_{14}}, {}^{1}|_{b_{15}}, {}^{0}|_{b_{16}}, {}^{0.9}|_{b_{17}}, {}^{0.3}|_{b_{18}}, {}^{0}|_{b_{19}}, {}^{0}|_{b_{20}} \right\}, \tag{5}$$

where elements of the set are in the form ${}^{\mu_{B_L}(b_i)}|_{b_i}$, $i = 1, \ldots, 20$.

According to the formula (3), a probability of a large ball drawing is obtained as follows:

$$p_F(B_L) = \frac{1}{20} \sum_{i=1}^{20} \mu_{B_L}(b_i) = \frac{1}{20} \cdot 9.8 = 0.49. \tag{6}$$

However, the result expresses the expected value of the randomly drawn ball membership degree to the fuzzy set B_L. Does this value really express a probability that a ball drawn at random is large? If e.g. the ball b_1 that is a "large ball" with a membership degree 0.5 will be drawn, are we able to say whether the event "a ball drawn at random is large" occurred?

5 Fuzzy probabilities of fuzzy events

Let us focus now on another way of expressing the probabilities of fuzzy events - by so called *fuzzy probabilities* whose membership functions are derived from the probabilities of α-cuts of fuzzy events. This means the resulting fuzzy probabilities reflect the fuzziness of the fuzzy events. The approach that will be described further was introduced by Yager [5] and by Talašová and Pavlačka [4]. In fact, the similar but not the same idea was proposed earlier by Zadeh [6].

Let $(\Omega, \mathcal{A}, p)$ is a given probability space, and let $\mathcal{A}_F$ denotes the family of all fuzzy events introduced in Section 4. The fuzzy probabilities of fuzzy events are assigned by a mapping $P_F : \mathcal{A}_F \to \mathcal{F}([0,1])$ that is defined in the following way: For any fuzzy event $A \in \mathcal{A}_F$, the membership function of the fuzzy probability $P_F(A)$ is given for all $\hat{p} \in [0,1]$ as follows:

$$\mu_{P_F(A)}(\hat{p}) = \begin{cases} \sup\{\alpha \in (0,1] \mid \hat{p} = p(A_\alpha)\}, & \text{if } \{\alpha \in (0,1] \mid \hat{p} = p(A_\alpha)\} \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

The membership function $\mu_{P_F(A)}$ is interpreted as a possibility distribution, i.e. the value $\mu_{P_F(A)}(\hat{p})$ means the degree of possibility that the probability of a fuzzy event A is equal to $\hat{p}$. For illustration, the membership function of the fuzzy probability $P_F(B_L)$ of the fuzzy event B_L that was defined in Example 1 is depicted in Fig. 1.

Let us examine now the properties of P_F. It can be easily seen from (7) that for any crisp event $A \in \mathcal{A}$, the fuzzy probability $P_F(A)$ "coincides" with the value $p(A)$; $\mu_{P_F(A)}(p(A)) = 1$, and $\mu_{P_F(A)}(\hat{p}) = 0$ for any $\hat{p} \neq p(A)$. Hence,

$$\mu_{P_F(\Omega)}(\hat{p}) = \begin{cases} 1, & \text{if } \hat{p} = 1, \\ 0, & \text{otherwise,} \end{cases}$$

i.e. the first property of a probability measure given by (1) can be observed also for P_F.

For showing the retaining of some other properties, it is convenient to introduce the following special operations $\oplus$ and $\ominus$ with fuzzy probabilities that represent an extension of the arithmetic operations sum

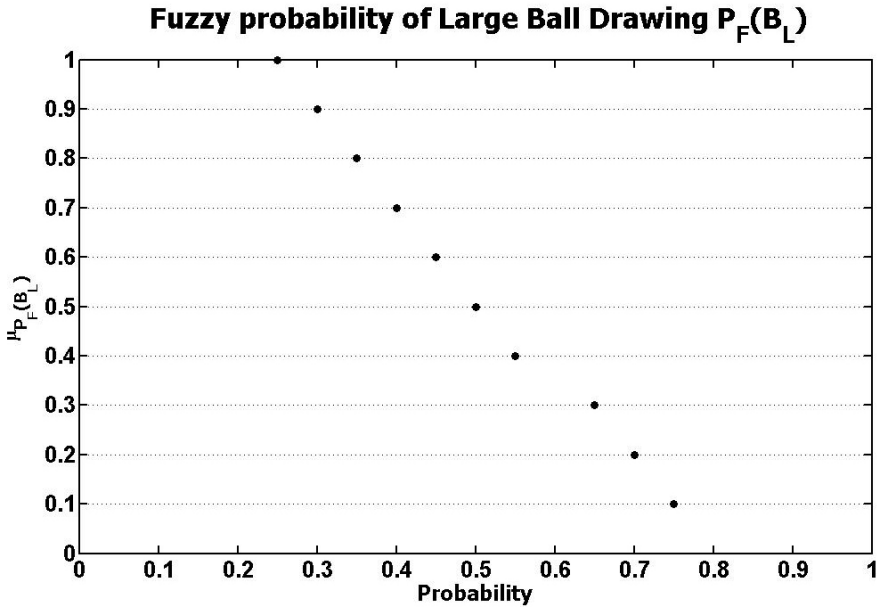


Figure 1 Membership function of the fuzzy probability $P_F(B_L)$.

and difference. The extension is not based on the well known *extension principle* proposed by Zadeh [7]. For any $A, B \in \mathcal{A}_F$, the membership functions of the fuzzy sets $P_F(A) \oplus P_F(B)$ and $P_F(A) \ominus P_F(B)$ are defined for all $\hat{p} \in \mathbb{R}$ in the following way:

$$\mu_{P_F(A)\oplus P_F(B)}(\hat{p}) = \begin{cases} \sup\{\alpha \in (0,1] \mid \hat{p} = p(A_\alpha) + p(B_\alpha)\}, & \text{if } \{\alpha \in (0,1] \mid \hat{p} = p(A_\alpha) + p(B_\alpha)\} \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases}$$

$$\mu_{P_F(A)\ominus P_F(B)}(\hat{p}) = \begin{cases} \sup\{\alpha \in (0,1] \mid \hat{p} = p(A_\alpha) - p(B_\alpha)\}, & \text{if } \{\alpha \in (0,1] \mid \hat{p} = p(A_\alpha) - p(B_\alpha)\} \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

The second fundamental property of a probability measure given by (2) can be observed for P_F in the following way:

Proposition 2. *For any $A_1, A_2, \ldots \in \mathcal{A}_F$ such that $A_i \cap A_j = \emptyset$ for any $i, j \in \mathbb{N}$, $i \neq j$:*

$$P_F\left(\bigcup_{i=1}^{\infty} A_i\right) = P_F(A_1) \oplus P_F(A_2) \oplus \ldots. \tag{8}$$

Proof. The assumptions imply that $A_{i\alpha} \cap A_{j\alpha} = \emptyset$ for any $i, j \in \mathbb{N}$, $i \neq j$ and for all $\alpha \in (0,1]$. Then, for any $\hat{p} \in [0,1]$:

$$\begin{aligned} \mu_{P_F(\bigcup_{i=1}^{\infty} A_i)}(\hat{p}) &= \begin{cases} \sup\{\alpha \in (0,1] \mid \hat{p} = p((\bigcup_{i=1}^{\infty} A_i)_\alpha)\}, & \text{if } \{\alpha \in (0,1] \mid \hat{p} = p((\bigcup_{i=1}^{\infty} A_i)_\alpha)\} \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \\ &= \begin{cases} \sup\{\alpha \in (0,1] \mid \hat{p} = p(\bigcup_{i=1}^{\infty} A_{i\alpha})\}, & \text{if } \{\alpha \in (0,1] \mid \hat{p} = p(\bigcup_{i=1}^{\infty} A_{i\alpha})\} \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \\ &= \begin{cases} \sup\{\alpha \in (0,1] \mid \hat{p} = \sum_{i=1}^{\infty} p(A_{i\alpha})\}, & \text{if } \{\alpha \in (0,1] \mid \hat{p} = \sum_{i=1}^{\infty} p(A_{i\alpha})\} \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \\ &= \mu_{P_F(A_1)\oplus P_F(A_2)\oplus\ldots}(\hat{p}). \end{aligned}$$

□

As for the other properties recalled in Section 2, only the first three can be observed for P_F: The fuzzy probability $P_F(\emptyset)$ coincides with zero, since $\mu_{P_F(\emptyset)}(0) = 1$, and $\mu_{P_F(\emptyset)}(\hat{p}) = 0$ for any $\hat{p} \neq 0$. If $A, B \in \mathcal{A}_F$,

$A \subseteq B$, then for all $\alpha \in (0,1]$, $\min\{\hat{p} \mid \hat{p} \in P_F(A)_\alpha\} \leq \min\{\hat{p} \mid \hat{p} \in P_F(B)_\alpha\}$, and $\max\{\hat{p} \mid \hat{p} \in P_F(A)_\alpha\} \leq \max\{\hat{p} \mid \hat{p} \in P_F(B)_\alpha\}$. These inequalities follow from the fact that $p(A_\alpha) \leq p(B_\alpha)$ for all $\alpha \in (0,1]$. Thus, we can conclude that $P_F(A) \leq P_F(B)$. For any $A, B \in \mathcal{A}_F$, $P_F(A \cup B) = P_F(A) \oplus P_F(B) \ominus P_F(A \cap B)$. It follows from the fact that for all $\alpha \in (0,1]$: $p((A \cup B)_\alpha) = p(A_\alpha \cup B_\alpha) = p(A_\alpha) + p(B_\alpha) - p(A_\alpha \cap B_\alpha) = p(A_\alpha) + p(B_\alpha) - p((A \cap B)_\alpha)$. The last two properties of a probability measure are not retained in the case of fuzzy events.

The fuzzy probability P_F whose membership function is given by (7) seems to be not fully appropriate to be used in practice. Its main advantage consists in the fact that it reflects the fuzziness of fuzzy events. Another advantage is that in contrast to p_F, P_F has the common probabilistic interpretation; from the fuzzy probability $P_F(A)$, we can see the probabilities of occurring the particular α-cuts of A which are crisp events. However, construction of fuzzy probabilities and subsequent calculation with them are not simple, particularly in the discrete case. Moreover, P_F does not retain some of the important properties of a probability measure.

6 Conclusion

We have examined two ways of expressing the probability of fuzzy events that are proposed in the literature. The first, most common way consists in expression the probability of a fuzzy event as the expected value of its membership function. We have shown that this approach has a nice mathematical properties, but lacks the common interpretation of a probability and does not reflect the fuzziness of fuzzy events. The second way is based on the idea that the probability of a fuzzy event should be also fuzzy. The membership function of the resulting fuzzy probability is derived from the probabilities of α-cuts of fuzzy events. However, the fuzzy probabilities do not have some important properties of a probability measure and the calculations with them could be awkward.

Thus, we have shown that none of the approaches gives an appropriate probability of a fuzzy event. The question "How should the probability of a fuzzy event be expressed?" is still an open problem.

Acknowledgements

The research is supported by the grant No. GA 14-02424S of the Grant Agency of the Czech Republic and by the grant IGA_PrF_2014028 Mathematical Models of the Internal Grant Agency of the Palacký University in Olomouc.

References

[1] Hisdal, E.: Are grades of membership probabilities? *Fuzzy Sets and Systems* **25** (3) (1988), 325–348.

[2] Kolmogorov, A.: *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Julius Springer, Berlin, 1933.

[3] Negoita, C. V. and Ralescu, D.: *Applications of Fuzzy Sets to System Analysis (1st ed.).* Birkhuser Verlag–Edituria Technica, Stuttgart, 1975.

[4] Talašová, J. and Pavlačka, O.: Fuzzy Probability Spaces and Their Applications in Decision Making. *Austrian Journal of Statistics* **35** (2&3) (2006), 347–356.

[5] Yager, R.R.: A note on probabilities of fuzzy events. *Information Sciences* **18** (2) (1979), 113-129.

[6] Zadeh, L. A.: Fuzzy Probabilities. *Information Processing & Management* **20** (1974), 363–372.

[7] Zadeh, L. A.: Fuzzy Sets. *Information and Control* **8** (1965), 338–353.

[8] Zadeh, L. A.: Probability Measures of Fuzzy Events. *Journal of Mathematical Analysis and Applications* **23** (1968), 421–427.

Parametric and Non-Parametric Calculation of Technical Efficiency: Application on the Czech Local Action Groups

Marie Pechrová[1]

Abstract. The aim of the paper is to calculate technical efficiency of the Czech Local Action Groups (LAGs) by parametric and non-parametric approaches and to compare the results. LAGs are actors concerned with rural development. They receive subsidies via Rural Development Program for the Czech Republic for years 2007–2013. Hence, the efficiency of the finances' usage should be examined. Their inputs and outputs are multiple. Firstly we used Stochastic Frontier Analysis (SFA) and estimated translog output distance function. The model was fitted by maximum likelihood. The true fixed effect assumed truncated-normal distribution of the inefficiency term. Heterogeneity among LAGs was explained in the mean of the inefficiency function. Secondly, the Data Envelopment Analysis (DEA) was used. Models of Charnes, Cooper and Rhodes assuming constant and of Banker, Charnes and Cooper for variable returns to scale were calculated. We used panel data of all 112 LAGs observed in 2012 and 2013. Average inefficiency was estimated by SFA at 21.28 % and efficiency at 89.29 %. The inefficiency decreased and efficiency increased overtime. On the other hand, CCR model predicted lower efficiency (56.04 %) as same as the BCC model (62.83 %). In both cases it decreased overtime.

Keywords: Local Action Groups, Stochastic Frontier Analysis, Output distance function, Data Envelopment Analysis, CCR and BCC models

JEL Classification: C23, C44
AMS Classification: 90C15

1 Introduction

"The endogenous, and, more recently, the neo-endogenous rural development concepts conceive the development as a process that arises from within the local areas" [6]. Local Action Groups (LAGs) were founded to enhance the development of rural areas where they operate. They work according to the LEADER. The key principle is that the local actors have the opportunity to influence the implementation of the local development strategy. LEADER is: (1) based on a specific geographic area, (2) applied in the bottom-up direction, (3) including partnerships between the public and private sectors, (4) integrating the various sectors in the region, (5) based on innovation, (6) forming a system of cooperation within and between territories and (7) based on a decentralized management and financing. "The local development strategy of LAGs is implemented in the form of projects, and it is possible to attract financial support from measures co-financed by the European Agricultural Fund for Rural Development (EAFRD) and the European Fisheries Fund (EFF) for its implementation in the period 2007–2013" [15]. Particularly, LAGs receive subsidies via Rural Development Program for the Czech Republic for years 2007–2013 (RDP), axis IV - LEADER. LAGs are subject of annual evaluation from the Ministry of Agriculture (MoA). We propose that LAGs should be evaluated according to the efficiency defined according to the European Commission [4] as the "assessment of the achieved effects in relation to the inputs (financial or administrative) mobilized". The paper is structured as follows. Firstly, the results of previous researches are presented. Than, the Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA) models are described. The results of calculations are presented, discussed and compared. Last chapter concludes.

The ideal of neo-endogenous development is that socio-economic well-being can be best (more effective and less costly) achieved by focusing on the needs and resources valorisation at a local level [12]. The effectiveness of the LAGs functioning concerns for example Kis [10]. He describes the framework in which the performance of the LAGs should be evaluated. Lošťák and Hudečková [7] evaluated to what extend LAGs inform about their activities. We build our analysis on the quantitative methods. There are two widely used to calculate efficiency: parametric SFA based on econometric modelling and non-parametric DEA based on linear programming.

The stochastic frontier analysis (SFA) model was independently developed by Aigner et al. [1] and Meeusen and van den Broeck [13] to describe the production function technology. The idea was to divide the error on two components: technical inefficiency error u_i and the white noise error component v_i. The v_i represents factors that

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management, department of economics, Kamýcká 129, 165 21 Prague 6, Czech Republic, pechrova@pef.czu.cz

might generate irrelevant noise in the data such as measurement error while u_i denotes unobserved inputs. The composed error term ε is sum of normally distributed variable (the idiosyncrasy) and the absolute value of a truncated normal, normal-exponential, or normal-gamma variable (the inefficiency). The models for panel data enables to consider heterogeneity in the mean of u_{it} and/or heteroscedasticity of v_{it} or/and u_{it}. Originally Fixed Effect Models and Random Effect Models were applied at SFA. Greene [5] stated the shortcomings of the two models and proposed "true" fixed and "true" random effects models.

The advantages on non-parametric method are that we do not have to presume any type of production function or any distribution of the inefficiency term. DEA also assumes each performance measure can be classified as either an input or an output. Hence, it might be more suitable for estimation of efficiency of units which are not classical firms (such are LAGs). The first DEA models were introduced by Charnes, Cooper and Rhodes [8]. So called CCR model assumed constant returns to scale. On the other hand, model introduced by Banker, Charnes and Cooper [2] consider in their BCC model variable returns to scale (decreasing, increasing or constant). As DEA assumes free disposability and convexity assumptions, it is further restricted by making an assumption on the shape of the convex hull or convex cone. A CCR model assumes constant returns to scale and convex cone. When the input increases by one the output increases proportionally.

2 Materials and Methods

Our sample consists of 224 observations (112 LAGs in 2012 and 2013). The average number of members was 41 and of inhabitants in the area 34170. The average LAG was allocated 1.24 mil. CZK for its functioning and its employees worked 6696.43 hours per year (which corresponds to 3 full time employees). It redistributed in average 3.21 mil. CZK in 1.5 calls for application. Its annual bonus in average was 3.66 mil. CZK. The data were obtained from State Agricultural Interventional Fund and from Ministry of Agriculture.

There are two approaches to the efficiency assessment – parametric and non-parametric widely utilized in researches. Some authors used both; hence, we also took the advantage of comparison the results. The input and output variables are the same for SFA and DEA. The subsidies for LAGs were provided from RDP under Axis IV. The finances for LAGs operation are mainly from measure IV.1.1 (input x_1). Second input (x_2) is the worked hours by employees per year. The outputs were based on the outcomes of the LAGs operation. LAGs are redistributing the finances to the final beneficiaries. According to the Guidelines for the Evaluation of Rural Development Programmes "the utilization of resources relates to the capacity of the decision group to assign the financial resources available to the project measures and can be measured as the proportion of the fund actually committed by the local group" [4]. Hence, we consider the amount of the redistributed subsidies from measure IV.1.2 as one output (y_1). The redistribution is performed via calls for submission of the projects. LAGs have to perform at least one call per year. The more they have, the more they prove that they are active and also the more inputs they consume. Therefore, the number of calls per year is the output y_2. Finally, the LAGs are evaluated and categorized according to their performance by the Ministry of Agriculture. Part of the subsidies for their functioning from measure IV.1.1 is annually allocated based also on this assessment. The more successful the LAG is, the more it is financed. This so called "bonus" is the third output y_3. For calculations Stata 11.2 software was used.

2.1 SFA – Output distance function

When many inputs produce many outputs, the distance function provides a functional characterization of the structure of the production technology. An advantage of the distance function is that neither the input distance nor the output distance function depends on any explicit behavioral assumptions such as cost minimization, revenue maximization, and profit maximization [11]. This justifies the usage of this approach in non-profit sector.

We have multiple inputs (n, j = 1... N) and multiple outputs (m, p = 1... M). We assume that each observed output of the LAG is scaled radially towards the boundary of the production function in order to operate on the production frontier. Then the output distance function takes an output–expanding approach to the measurement of the distance of a producer to this boundary. "It gives the minimum amount by which an output vector can be deflated by factor θ and still remain producible with a given input vector" [14]. We estimated the translog form of a production function. Applying the assumptions of homogeneity of degree 1, monotonicity assumptions and curvature requirements on the translog transformation function, we got (1). We normalized the outputs by y_3.

$$-\ln y_m = \alpha_0 + \sum_{p=1}^{M-1}\alpha_p \ln\left(\frac{y_p}{y_m}\right) + \sum_{n=1}^{N}\beta_n \ln(x_n) + \frac{1}{2}\sum_{p=1}^{M-1}\sum_{o=1}^{M-1}\mathbf{A}_{po}\ln\left(\frac{y_p}{y_m}\right)\ln\left(\frac{y_o}{y_m}\right) + \frac{1}{2}\sum_{n=1}^{N}\sum_{j=1}^{N}\mathbf{B}_{nj}\ln(x_n)\ln(x_j) + \frac{1}{2}\sum_{n=1}^{N}\sum_{p=1}^{M-1}\mathbf{C}_{np}\ln(x_n)\ln\left(\frac{y_p}{y_m}\right) + u_{it} - v_{it} \quad (1)$$

where α_0 is constant, α_m (m = 1, 2, 3) are parameters related to the outputs, β_n (n = 1, 2) are the parameters for inputs, $\mathbf{A}_{mp}$ is the matrix of the parameters related to the cross-products of the outputs, and $\mathbf{B}_{nj}$ is the matrix of the parameters for cross-products of the inputs, and $\mathbf{C}_{nm}$ denotes the parameters related to the cross-products of

inputs and outputs. We did not include time variable as we have observations only for two years. u_{it} represents the inefficiency term and v_{it} the stochastic term. We assume that u_{it} follows truncated normal distribution $N^{+}(\mu_{u_{it}};\sigma^2_{u_{it}})$ and v_{it} is independently identically distributed $N(0;\sigma^2_{v_{it}})$. We suppose heterogeneity among LAGs. Therefore, we included two explanatory variables into the function of the mean of the inefficiency term: z_1 – the number of members of the LAG and z_2 – the number of inhabitants. Also a function the variances or inefficiency term and stochastic term (2) were specified.

$$\mu_{u_{it}} = \exp(\delta_0 + \delta_1 z_1 + \delta_2 z_2); \sigma^2_{u_{it}} = \exp(\omega_0); \sigma^2_{v_{it}} = \exp(\gamma_0) \tag{2}$$

The inefficiency term, u_{it} can be interpreted as the percentage deviation of observed performance, y_{it} from the firm's own frontier performance [5]. In order to satisfy the monotonicity assumption, the signs of inputs in output distance function should be negative and of outputs positive. Parameters can be interpreted as elasticity of outputs at sample mean. We fitted True fixed effect (TFE) model as specified by Greene [5] by maximum likelihood method in the form $y_{it} = \alpha_i + \boldsymbol{\beta}^T \boldsymbol{x}_{it} + v_{it} - u_{it}$, where α_i is the group specific constant and u_{it} varies in time. This form also allows the heterogeneity term to be correlated with the explanatory variables. The inefficiency (1-Exp(u_{it})) was calculated based on the expected value of u_{it} knowing ε_{it} as $\hat{u}_{it} = E[u_{it}|\varepsilon_{it}]$, where $\sigma = \sqrt{(\sigma_v^2 + \sigma_u^2)}$, $\lambda = \sigma_u / \sigma_v$, $a_{it} = \varepsilon_{it}\lambda / \sigma_v$ for production and $a_{it} = -\varepsilon_{it}\lambda / \sigma_v$ for cost function. $\phi(a_{it})$ is standard normal density evaluated at a_{it} and $\Phi(a_{it})$ is the standard normal CDF (integral from -∞ to a_{it}) evaluated at a_{it}.

2.2 Input oriented DEA model

We considered constant (CCR model) and variable (BCC model). As the LAGs inputs (public finances) are given, we assume that the outputs generated shall be as high as possible. We choose input oriented DEA model. Assuming constant returns to scale, we used CCR model to maximize the efficiency of the assessed LAG expressed as the division of weighted outputs and weighted inputs. Constrains state that efficiency of other LAGs is less or equal 1. For each LAG_q (q = 1... 112) we weight each input x by v_j, j = 1,... n and gain virtual input and weight each output y by u_i, i = 1, ... m and get virtual output.

CCR DEA model computes weighs of inputs and outputs for each LAG, presuming that the maximal score for other LAGs is 1. We use Charnes-Cooper transformation (see e.g. [9]) and formulate primal CCR model (3):

$$\begin{aligned} \max \quad & z = \sum_{i=1}^{m} u_i y_{iq}, \\ \text{s. t.} \quad & \sum_{i=1}^{m} u_i y_{ik} \le \sum_{j=1}^{n} v_j x_{jk}, && k = 1, 2, ..., 112, \\ & \sum_{j=1}^{n} v_j x_{jq} = 1, \\ & u_i \ge \varepsilon, && i = 1, 2, ..., m, \\ & v_j \ge \varepsilon, && j = 1, 2, ..., n \end{aligned} \tag{3}$$

where z is efficiency of unit LAG_q, x_{jk} is value of j^{-th} input for unit LAG_k, y_{ik} is value of i^{-th} output for unit LAG_k. There are n inputs, m outputs and *112* LAGs, ε is infinitesimal constant ensuring that all weights of inputs and outputs will be positive and included in a model. The value is usually 10^{-8}. Again, v_i represents weights for inputs (i = 1, 2, ... m) and u_i are weights for outputs (i = 1, 2, ... n). For calculations and interpretation, a dual model in matrix form is used. A computational form of this model is (4)

$$\begin{aligned} \min \quad & z = \theta_q - \varepsilon(\boldsymbol{e}^T \boldsymbol{s}^+ + \boldsymbol{e}^T \boldsymbol{s}^-), \\ \text{s. t.} \quad & \boldsymbol{X}\lambda + \boldsymbol{s}^- = \theta \boldsymbol{x}_q, \\ & \boldsymbol{Y}\lambda - \boldsymbol{s}^+ = \boldsymbol{y}_q, \\ & \lambda, \boldsymbol{s}^+, \boldsymbol{s}^- \ge \boldsymbol{0}. \end{aligned} \tag{4}$$

where θ_q is the efficiency of LAG_q, $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)$, $\lambda \ge 0$ - λ are weights for particular LAG, $\mathbf{s}^+$ and $\mathbf{s}^-$ are slacked variables in restrictions for inputs and outputs, $\mathbf{e}^T$ = (1, 1, ... 1) and ε is infinitesimal constant (10^{-8}). θ_q can be interpreted as needed input reduction for the achievement of the efficiency frontier. When it is equal to 1, the LAG is 100 % efficient. The model tries to find a virtual LAGs for LAG_q, which are linear combination of inputs and outputs of other units in a dataset and which inputs and outputs are not worse of the evaluated unit. Optimal values of slack variables (s^{*+}, i = 1, ... m and s^{*-} i = 1, ... n) are equal to zero.

BCC model assumes variable returns to scale. Its formulation is similar to CCR, but includes one variable taking arbitrary value, which accounts for the convexity. We can write primal BCC input oriented model as (5).

$$\begin{aligned} \max \quad & z = \sum_{i=1}^{m} u_i y_{iq} + \mu, \\ \text{s. t.} \quad & \sum_{i=1}^{m} u_i y_{ik} + \mu \le \sum_{j=1}^{n} v_j x_{jk}, && k = 1, 2, ..., 112 \\ & \sum_{j=1}^{n} v_j x_{jq} = 1, \\ & u_i \ge \varepsilon, && i = 1, 2, ..., m, \\ & v_j \ge \varepsilon, && j = 1, 2, ..., n \\ & \mu - \text{any} \end{aligned} \tag{5}$$

where μ can take any value. When it is lower or equal to 0, there are non-increasing returns to scale and when higher or equal to zero, than non-decreasing ones. We can rewrite the model in matrix form (6), where $\mathbf{e}^T\lambda = 1$ is the convexity condition.

$$\begin{aligned} \min \quad & z = \theta_q - \varepsilon(\boldsymbol{e}^T\boldsymbol{s}^+ + \boldsymbol{e}^T\boldsymbol{s}^-), \\ \text{s. t.} \quad & \mathbf{X}\lambda + \mathbf{s}^- = \theta\mathbf{x}_q, \\ & \mathbf{Y}\lambda - \mathbf{s}^+ = \mathbf{y}_\mathbf{q}, \\ & \mathbf{e}^T\lambda = 1, \\ & \lambda, \mathbf{s}^+, \mathbf{s}^- \ge \mathbf{0}. \end{aligned} \tag{6}$$

3 Results

Firstly, a parametric approach was applied on panel with 224 observations. Translog output distance function was estimated as "true" fixed effect model [5] assuming truncated-normal distribution of inefficiency term. The results are displayed at Table 1. Wald $\chi^2 = 1.94.10^6$ with p-value = 0.0000 revealed that the model as a whole is statistically significant. The value of Log likelihood was = 100.0276. The signs of inputs are negative and of outputs positive according to expectations. The variables were included in logarithms. Hence, the parameters can be interpreted as elasticity of the outputs at a sample mean. A marginal increase in outputs ceteris paribus implies an improvement in the efficiency, i.e. a decrease of the distance to the stochastic frontier. Negative sign of inputs mean positive effect on technical efficiency. The function of the inefficiency mean shows that the increase of the LAG's size (in terms of the number of members and inhabitants per LAG) statistically significantly decreases the inefficiency. This suggests that bigger LAGs may have the advantage from having more resources. Not only more finances, but also more employees can bring the LAG closer to the efficiency frontier.

Average inefficiency was 21.28 %, while the LAGs were more inefficient in year 2012. The inefficiency decreased the year after. The average efficiency of LAGs was 86.29 %. The most inefficient LAGs had legal form civic association (22.79 % in average); while the most efficient were professional associations of legal persons (96.36 % in average). Regarding the region, the most inefficient LAGs came from South-East NUTS II (24.95 %). Their average efficiency was only 86.90 %. Also LAGs from South-West should improve by 12.24 % to be 100 % efficient. The most efficient was Central Moravian and Moravian-Silesian regions (92.40 % and 91.48 % in average, resp.). We also tested, whether the differences among regions are statistically significantly different. Firstly, the Shapiro-Wilk W test revealed that the distribution of the data is not normal. Hence, we used non-parametric analogy of ANOVA - Kruskall-Wallis test for equality of medians. Null hypothesis for Kruskall –Walis test assumes that the (in)efficiency and in all regions have similar medians. Because the calculated probability p-value was 0.8243 for inefficiency and 0.8244 for efficiency, we could not reject null hypothesis. Despite that there are not differences among regions in in/efficiency, there were found statistically insignificant. We also examined further the relation between the size of the LAG and its efficiency. Spearman rank correlation coefficient shows that there is weak positive relation between number of members of the LAG and technical efficiency. The more members the LAG has, the more efficient it is. Similarly, the more inhabitants live in the area, the more efficient the LAG tends to be. When the LAG has above average members (over 41), it tends to be less inefficient (25.32% average inefficiency for "small" and 15.48% for "large"). Also the average efficiency is higher in later case (90.80 % in comparison with 88.24 %). The same pattern is true when we measure the size of the LAG in terms of the number of inhabitants living in the area. When the LAG is large (has above 34170 inhabitants), it is less inefficient (only from 9.91 % in average) than small one (33.02 % in average). Hence, it is not surprising, that also efficiency is higher in a group of larger LAGS (91.69 %) than in a group of smaller ones (86.51 %).Using Mann-Whitney test, we found out that the differences in inefficiency between large and small

LAGs (in terms of the number of members) are not statistically significant (z =2.01**). The same is true for efficiency (z =-2.01**). Regarding the size measured as the number of inhabitants, we found that the differences in inefficiency (z = 3.15***) and efficiency (z=-3.14**) are not statistically significant. This might be due to the fact, that there is higher human capital potential when there are more people involved in the SPL implementation. Regarding the relation of the LAG and subsidies we found the correlation among (in)efficiency and subsidies is relatively weak. Higher is for subsidies that are redistributed by the LAG (Spearman rank correlation coefficient between efficiency and subsidies from IV 1.2 is 0.1758***). Similarly, the more subsidies the LAG receives, the higher is efficiency (0.1414**). Same coefficients but with opposite sign are true for inefficiency.

Variable	Coef.	Variable	Coef.
Frontier			
α_1 [ln(y_1/y_3)]	0.004061 (0.000656)***	β_1 [ln(x_1)]	-.048258 (N/A)
α_2 [ln(y_2/y_3)]	0.094500 (0.000697)***	β_2 [ln(x_2)]	-.198989 (N/A)
a_1 [ln(y_1/y_3).ln(y_1/y_3)]	0.000152 (0.000014)***	b_1 [ln(x_1)ln(x_1)]	-0.010590 (.0000434)***
a_2 [ln(y_1/y_3).ln(y_2/y_3)]	-0.000211 (0.000031)***	b_2 [ln(x_1)ln(x_2)]	-0.004816 (.0000601)***
a_3 [ln(y_2/y_3).ln(y_2/y_3)]	-0.009874 (0.000043)***	b_3 [ln(x_2)ln(x_2)]	-0.027082 (.0000356)***
c_1 [ln(x_1).ln(y_1/y_3)]	-0.000584 (.000046)***	c_3 [ln(x_2).ln(y_1/y_3)]	0.000481 (0.000050)***
c_2 [ln(x_1).ln(y_2/y_3)]	0.006094 (0.000038)***	c_4 [ln(x_2).ln(y_2/y_3)]	0.015178 (0.000054)***
Mean of inefficiency (μ_u)		**Variance of inefficiency (σ^2_u)**	
δ_1 [z_1]	-0.605678 (0.266964)**	ω_0 [const.]	2.249302 (0.354798)***
δ_2 [z_2]	-0.000979 (0.000390)**	**Variance of stochastic term (σ^2_v)**	
δ_0 [const.]	5.067866 (7.491456)	γ_0 [const.]	-21.085670 (6.060632)***

Table 1 Estimation results of output distance function (TFE model); **Source** Own elaboration; **Note** Statistical significance is marked as follows: * on α = 0.1, ** on α = 0.05 and *** on α = 0.01 level.

Secondly, the efficiency of the LAGs was estimated by DEA using CCR and BCC model. The average efficiency in the first model was calculated at 64.23 % in 2012 and 47.95 % in 2013. In BCC model, the efficiency also decreased (71.32 % in 2012 and 54.33 % in 2013). It is natural, that the efficiency had higher estimated values in BCC than in CCR model. While there were only 6 (2012) and 5 (2013) LAGs 100 % efficient in the CCR, BCC predicted 12 (2012) and 11 (2013) LAGs. It is due to the fact that the data envelope in CCR is conical, whereas in BCC in changes to convex. The number of efficient units decreased. This is in contrast to the results calculated by SFA, where the efficiency increased overtime. The estimates of the technical efficiency differ because of the different functional forms assumption and variation of the values in a sample. While SFA estimated translog function, CCR and BCC did not assume any. In line with De Borger and Kerstens [3] we observe that the shape of the efficiency distribution may be affected by the use of different reference technologies and by implied rankings of individual observations. The results strongly correlate. Spearman rank coefficient was 0.8448*** for 2012 and 0.8910*** for 2013. The highest correlation was found between results from DEA CCR and SFA for year 2013 (0.2966***). Also for year 2012 there was a weak correlation (0.2856***). Despite that the efficiency of BCC model was closer to the efficiency from SFA, the correlation between the results is weak (0.0792 in 2012 and 0.2507***). De Borger and Kerstens [3] also found high correlation between DEA and SFA efficiencies. The results of both approaches are at Table 2.

Variable	Year	Mean	Std. Dev.	Min	Max
SFA	2012	0.862190	0.228346	0.004588	0.999906
DEA CCR		0.642294	0.153015	0.359825	1.000000
DEA BCC		0.713239	0.153534	0.450754	1.000000
SFA	2013	0.923615	0.132313	0.479216	0.999905
DEA CCR		0.479525	0.193366	0.127011	1.000000
DEA BCC		0.543297	0.220620	0.131333	1.000000

Table 2 Comparison of technical efficiencies calculated by SFA and DEA; **Source** Own elaboration

4 Conclusion

The aim of the paper was to calculate technical efficiency of the Local Action Groups (LAGs) in the Czech Republic by parametric and non-parametric approaches and to compare the results with each other. In year 2012, the average efficiency was calculated at the level of 86.22 % by SFA, 64.23 % by DEA CCR model and 71.32 % by BCC model. According to SFA, the efficiency increased in 2013 (92.36 %), but DEA suggested opposite development (47.85 % in CCR and 54.22 % in BCC). The second conclusion is more probable, as the year 2013 was the last of the EU's programming period 2017–2013. There was an uncertain situation about the continuation of the subsidization of the LAGs and about the framework of their functioning in the next period. Hence, this might have influenced the efficiency of LAGs operations. On the other hand, our panel is too short to conclude on the efficiency development. Regarding the spatial distribution, the most efficient was Central Moravian and Moravian-Silesian NUTS II regions, but the differences among regions were not statistically significant.

From the factors, which can influence the mean of the inefficiency, we examined so far only the size of the LAG. We concluded that the larger is LAG (in terms of the number of members and in terms of the number of inhabitants), the less inefficient it is. Hence, the challenge for future research is to examine in detail the determinants of LAGs' inefficiency.

Acknowledgements

Supported by the grant No. 11110/1312/3179 – "Vliv dotací z Programu rozvoje venkova na technickou efektivnost příjemců" of the Internal Grant Agency of the Faculty of Economics and Management, Czech University of Life Sciences Prague.

References

[1] Aigner, D., Lovell, C. A. K., Schmidt, P.: Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* **6** (1977), 21–37.

[2] Banker, R. D., Charnes, A. W., Cooper, W. W., Swarts, M. J. and Thomas, D.: An introduction to data envelopment analysis with some of its model and their uses. *Research in Governmental and Nonprofit Accounting* **5**(1989), 125-163.

[3] De Borger, B. and Kerstens, K.: Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches. *Regional Science and Urban Economics* **26** (1996), 145-170.

[4] European Commission: *Guidelines for the evaluation of LEADER+ programmes supported by SAPARD*. Directorate General for Agriculture, European Commission, Brussels, 2001.

[5] Greene, W.: Fixed and Random Effects in Stochastic Frontier Models. 2002. [Online]. Available: http://people.stern.nyu.edu/wgreene/fixedandrandomeffects.pdf [Accessed: 01 Apr. 2013].

[6] High, C., Nemes, G.: Social Learning in LEADER Exogenous, Endogenous and Hybrid Evaluation in Rural Development, *Sociologia Ruralis* **47** (2007), 103-119.

[7] Hudečková, H., Lošťák, M.: Structural policy EU and its implementation in CR – application of Leader approach. In: *Financial support of agriculture and regional development – selected aspects*. Powerprint, Prague, 2011, 99–119.

[8] Charnes, A. W., Cooper, W. W., Rhodes, E.: (1978). Measuring efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429 – 444.

[9] Jablonský, J. and Dlouhý, M.: *Modely hodnocení efektivnosti produkčních jednotek*. Professional Publishing, Prague, 2004.

[10] Kis, K.: The experiences of Implementing LEADER Approach, with Special Attention to Effectiveness, Efficiency and Sustainability. Studia Universitatis Babes Bolyai, Oeconomica **2** (2012), 34-48.

[11] Kumbhakar S.C., Lovell C. A. K.: *Stochastic Frontier Analysis*. University Press, Cambridge, 2000.

[12] Lopolito, A., Nardone, G., Sisto, R.: Towards a Comprehensive Evaluation of Local Action Groups in LEADER Programmes. *New Medit* **1** (2011), 43–49.

[13] Meeusen, W., van den Broeck, J.: Technical efficiency and dimension of the firm: Some results on the use of frontier production functions. *Empirical Economics* **2** (1977), 109-122.

[14] Ogundari, K. and Brümmerb, B.: Estimating Technical Efficiency, Input substitution and complementary effects using Output Distance Function: A study of Cassava production in Nigeria. In: *Proceedings of the Joint 3rd African Association of Agricultural Economists (AAAE) and 48th Agricultural Economists Association of South Africa (AEASA) Conference*. AAAE & AEASA, Cape Town, 2010, 1-28.

[15] Silina, L., Zvaigzne, A., Dimanta, Z.: Implementation Evaluation of Strategies of Local Action Groups in Latvia. European Integration Studies **6** (2012), 69-76.

Randomized heuristics for capacitated arc routing problem

Jan Pelikán [1]

Abstract. This paper addresses the problem of providing services or goods to the edges of the graph which models the roads or paths of the communication network. The task is to search for a closed path which contains either all edges or a set of desired edges, where the length of the path is minimized. There are many modifications of the problem: postman problem, rural postman problem, windy postman problem, capacitated postman problem. Those tasks have a number of practical applications, such as the optimization of municipal waste collection, route optimization problem for cleaning city streets, reading energy meters, delivery of mail, school bus routes. This article outlines the tasks, models and solution methods.

Keywords: arc routing, capacitated postman problem, randomized heuristics, case study

JEL classification: C44
AMS classification: 90C15

1 Introduction

Methods of graph theory are used for modelling and solving logistical problems. Travelling salesman problem and vehicle routing problem solve the distribution of goods from a depot to the points of a communication network based on the demand for goods in those points. The goal is the minimization of transportation costs. A graph is the model of the communication network, one node of the graph is the depot and the other are points of the communication network. The edges of the graph are connections between points of the network or the shortest paths between two points of the network. This paper addresses the problem of providing services or goods to the edges of the graph which model the roads or paths of the communication network. The task is then the problem of searching for a closed path which contains either all edges or a set of desired edges, where the length of the path is minimized. The oldest and best-known problem of this type is searching for an Euler closed path in the graph. A generalization is the postman problem with a number of modifications: rural postman problem, windy postman problem, capacitated postman problem (caled arc routing problem CARP). These tasks have a number of practical applications, such as the optimization of municipal waste collection, route optimization problem for cleaning city streets, reading energy meters, delivery of mail, school bus routes.

In the paper the mathematical model of CARP is introduced and randomized heuristic method proposed. Numerical experiments with the heuristics are presented.

2 The model of the CARP on undirected graph

Let $G = \{V, E\}$ be an undirected graph, where $V = \{1, 2, \ldots, n\}$ is a set of n nodes and E is a set of nondirect arcs. Let us define $R \subset E$ as the set of obligatory arcs. For each arc $(i, j) \in E$ the length c_{ij} is given, for arc $(i, j) \in R$ the service requirement q_{ij} is known. Depot is situated in node 1, where m vehicles with capacity W are available for arcs service.

Let integer variables x_{ij}^k represent a number of times arc $(i, j) \in E$ is traversed by vehicle k (in any direction). Binary variable y_{ij}^k is equal to 1 if arc $(i, j) \in R$ is serviced on route k, i.e. service requirement q_{ij} is ensured by the vehicle k, 0 otherwise. Variables y_{ij}^k and x_{ij}^k are defined only for $i < j$.

[1] University of Economics Prague, W. Churchill sq. 4, 13067 Praha 3, Czech Republic, pelikan@vse.cz

Mathematical model of CARP:

$$\sum_{k=1}^{m} \sum_{(i,j)\in E} c_{i,j} x_{ij}^k \to min. \tag{1}$$

$$y_{ij}^k \leq x_{ij}^k, \quad \text{for} \quad (i,j) \in R, \quad k = 1, 2, \ldots, m. \tag{2}$$

$$\sum_{k=1}^{m} y_{ij}^k = 1, \quad \text{for} \quad (i,j) \in R. \tag{3}$$

$$\sum_{(i,j)\in R} q_{ij} y_{ij}^k \leq W, \quad \text{for} \quad k = 1, 2, \ldots, m. \tag{4}$$

$$\sum_{(i,j)\in\delta(i)} x_{ij}^k = 2f_i^k, \quad \text{for} \quad k = 1, 2, \ldots, m, i = 1, 2, \ldots, n \tag{5}$$

$$\sum_{(i,j)\in\delta(S)} x_{ij}^k \geq \frac{1}{M} \sum_{(i,j)\in E(S)} x_{ij}^k, \quad \text{for} \quad k = 1, 2, \ldots, m, S \subset V - \{1\}, S \neq \emptyset \tag{6}$$

The objective function (1) is total length of all arcs multiplied by it's multiplicity x_{ij}^k. If arc (i, j) is serviced on route k it must lie on this route, therefore (2) is valid. Because each arc must be serviced exactly once, i.e. only on one route, the equations (3) must hold. The sum of requirements q_{ij} serviced on route k must not exceed the vehicle capacity W, which is assured by capacity conditions (4).

Next, two other conditions have be added to the model:
a) parity conditions: all nodes have even degree (a number of arcs incident with the node), which is assured by the equations (5) ($\delta(i)$ is a family of arcs incident with the node i),
b) connectivity conditions: each node lying on the cycle has to be connected with node 1 with arcs on this cycle, the inequalities (6) assure these conditions. Set $\delta(S)$ is the set of all arcs connecting set S and $V - S$, and set $E(S)$ is the set of all arcs incident with only nodes from S.

Comment:

Connectivity conditions (6) can be formulated alternatively as
$\sum_{(i,j)\in\delta(S)} x_{ij}^k \geq \frac{1}{|R|} \sum_{(i,j)\in R} y_{ij}^k, \quad \text{for} \quad k = 1, 2, \ldots, m, S \subset V - \{1\}, S \neq \emptyset.$

In [1] the connectivity conditions are in the following form
$\sum_{(i,j)\in\delta(S)} x_{ij}^k \geq 2y_{ij}^k, \quad \text{for} \quad k = 1, 2, \ldots, m, S \subset V - \{1\}, S \neq \emptyset, i, j \in S.$

The model is solved by row generation manner: at first it is solved model (1)-(5), i.e. without conditions (6). If there are illegal subtours in the optimal solution, the inequality (6) for each subtour is formulated and added to model and resolved again. The method is finished if no illegal subtour exists in the optimal solution.

Unfortunately it can be a hudge number of constraints (6) in the model and then the model may not be solvable for a large graph. For this reason and for NP hardness of CARP heuristic methods are proposed and used in practice.

A disadvange of heuristics is that it does not provde the optimal solution even though it uses a rational rules for route creating. The proposed randomized heuristics uses this rules only with certain probability and route creating is repeated many times and the best solution obtained is the result of the heuristics.

3 The randomized heuristic algorithm

In literature authors are inspired by heuristics developed for vehicle routing problems and especially for travelling salesman problem: insert methods, savings algorithm, greedy approach, etc. Subsequently, improving methods (e.g. exchange algorithm, tabu search and other metaheuristics [3] [4]) can be applied to decrease cost of the generated solution. Proposed heuristics often creates the solution on the base of deterministic strategy for selecting arcs into the route. Unfortunately, such rules can lead to the solution distant from the optimum. We propose the algorithm which does not follow this approach, but allows the random selection of arcs which can result in more effective solution. In this Monte Carlo simulation, the final result is the best one obtained after a large number of generation runs.

We will use the following notation in proposed randomized heuristics:

Notation.

$P = \{(ij); (i,j)$ is the shortest path from node i to node $j\}$
d_{ij} is the length of the path $(i,j) \in P$
$\mathbf{v} = (v_1, v_2, \ldots, v_s)$, where $v_1 = v_s = 1$ and $(v_i, v_{i+1}) \in E \bigcup P, i = 1, 2, \ldots, s$
F objective of the solution $\mathbf{v}$
F^* objective of the best solution found $\mathbf{v}^*$
kd_{ij} is equal 0 if service of arc (i,j) is not yet covered, 1 otherwise
t index of the last node v_t included in the solution $\mathbf{v}$
Q_t the total sum of requirements serviced on current route after visiting node v_t
$Nsim$ number of solutions randomly generated by the heuristics.

Algorithm.

Step 0: Put $ksim := 1$, $F^* := \infty$

Step 1: (initialization of new solution v)
Put: $t := 1$, $v_1 = 1$, $F := 0$, $Q_t := 0$, $n_R := |R|$,
$n_1 := n_2 := 0$,
$kd_{ij} := 0, (i,j) \in R$

Step 2: (create the set R_t of obligatory arcs incident with node v_t)
if $n_R = 0$ then go to Step 6
$R_t := \{(v_t, j) \in R|, kd_{v_t,j} = 0, Q_t + q_{v_t,j} \leq W\}$,
$n_1 := |R_t|$
if $n_1 = 0$ the go to Step 4

Step 3: (choose an arc from set R_t)
select an arc (v_t, j) from set R_t according to some random choise
$v_{t+1} := j, F := F + c_{v_t,v_{t+1}}; Q_{t+1} := Q_t + q_{v_t,v_{t+1}}$;
$t := t + 1$; $n_R := n_R - 1$;
if $n_R = 0$ then go to Step 6
if $F \geq F^*$ then go to Step 1
if $v_t = 1$ then $Q_t := 0$
go to Step 2

Step 4: (create set P_t of paths from node v_t)
$P_t := \{(v_t, j) \in P | \exists (j,l) \in R; kd_{jl} = 0; Q_t + q_{jl} \leq W\}$
$n_2 := |P_t|$;
if $n_2 = 0$ then go to Step 6 else go to Step 5

Step 5: (choose a path from set P_t)
select path (v_t, j) from set P_t according to minimal value of $d_{v_t,j}$
$v_{t+1} := j; F := F + c_{v_t,v_{t+1}}; t := t + 1$
if $F \geq F^*$ then go to Step 1
if $v_t = 1$ then $Q_t := 0$
go to Step 2

Step 6: (return to the depot)
if $v_t \neq 1$ then $v_{t+1} := 1; Q_{t+1} := 0; F := F + d_{v_t,1}; t := t + 1$
if $F \geq F^*$ then go to Step 1
if $n_R = 0$ then go to Step 7 else go to Step 2

Step 7: if $F < F^*$ then $F^* := F; \mathbf{v}^* := \mathbf{v}$;
$ksim := ksim + 1$;
if $ksim < Nsim$ then go to Step 1 else Stop

Notes to the algorithm:

In Step 3, obligatory arc (v_t, j) is chosen from set R_t with the use of pseudorandom generator of values following given probability distributions:
1) Uniform, each arc is chosen with the probability value of $1/n_1$.
2) Non-uniform, arcs with greater service requirement have higher probability to be chosen.
3) Non-uniform, arcs with closer node j have higher probability to be chosen.

The possibility 3) is used if the sum of service requirements is near to the vehicle capacity.

Numerical experiements with randomized heuristics for CARP showed that the choice of edges in the Step 3 with uniform probability distribution (version 1)) caused a slower decrease of the cost function compared with non-uniform probability distribution (version 2) and 3)).

Results.

Suggested heuristics was approved on the set of testing instances also used in (Longo at al. 2006 [5] [6]). Obtained results are shown in Table 3 with the following information in columns: codename of the instance, number of nodes, number of arcs, number of serviced arcs, cost of the lower bound presented in literature (LB) published in (Longo at al. 2006 [5] [6]), cost of the best solution obtained using suggested heuristics after 1000 generation runs (r1000) and 10000 runs (r10000) followed by percentage deviation from LB.

Conclusion.

The randomized heuristics for CARP is proposed and tested on database of instances and on a case study in the paper. Obtained solution may then be used as starting solution for improving metaheuristics.

The randomized heuristics was also used to solve the case study of garbage collection in Podebrady city. It contains 142 obligatory arcs, 4 non-obligatory arcs and 101 nodes representing crossroads. The optimal solution of the problem using of the model (1)-(5) without cyclical restrictions (6) was not obtained because of huge size of the problem. Interrupting the CPLEX calculation after 24 hours, we obtained the estimation of 26 728 meters, what is also the lower bound estimation of CARP. Because we did not reach the optimal solution we used the proposed heuristics. After 5000 generation runs we obtained the objective value of 35 330 meters that is by 16 % lower than the existing distance value given by the real garbage collection company.

Acknowledgements

This publication was supported by the grant no. P403/12/1947 of the Grant Agency of the Czech Republic, and the grant no. FIS 11/2013 of the Internal Grant Agency of the University of Economics, Prague.

	$\mid V \mid$	$\mid E \mid$	$\mid R \mid$	LB	r1000	%LB	r10000	%LB
e1-a	77	98	51	3548	4104	15,7	3913	10,3
e1-b	77	98	51	4468	4975	11,3	4956	10,9
e1-c	77	98	51	5542	6762	22,0	6596	19,0
e2-a	77	98	72	5011	5655	12,9	5630	12,4
e2-b	77	98	72	6280	7298	16,2	7233	15,2
e2-c	77	98	72	8234	9480	15,1	9476	15,1
e3-a	77	98	87	5898	6725	14,0	6593	11,8
e3-b	77	98	87	7697	8848	15,0	8671	12,7
e3-c	77	98	87	10163	11669	14,8	11551	13,7
e4-a	77	98	98	6395	7401	15,7	7153	11,9
e4-b	77	98	98	8884	10093	13,6	10017	12,8
e4-c	77	98	98	11427	13400	17,3	13272	16,1
s1-a	140	190	75	5014	5941	18,5	5941	18,5
s1-b	140	190	75	6379	8114	27,2	8110	27,1
s1-c	140	190	75	8480	9621	13,5	9460	11,6
s2-a	140	190	147	9824	11618	18,3	11330	15,3
s2-b	140	190	147	12968	14921	15,1	14914	15,0
s2-c	140	190	147	16353	19116	16,9	19116	16,9
s3-a	140	190	159	10143	12432	22,6	12255	20,8
s3-b	140	190	159	13616	16122	18,4	15900	16,8
s3-c	140	190	159	17100	20262	18,5	20172	18,0
s4-a	140	190	190	12143	14644	20,6	14510	19,5
s4-b	140	190	190	16093	19077	18,5	18845	17,1
s4-c	140	190	190	20375	24105	18,3	23631	16,0

Table 1 Results

References

[1] Eiselt, H. A., Gendreau, M. and Laporte, G. *Arc Routing Problems, Part II: The Rural Postman problem.* Operations Research. Vol. 43, No. 3, 1995, pp. 399-414.

[2] Belenguer, J., Benavent, E., Labadi, N., Princ, Ch., Reghioui, M.:*Split Delivery Capacitated Arc-Routing Problem: Lower Bound and Metaheuristic* . In Transportation Science, Vol. 44, No. 2, 2010 pp. 206-220.

[3] Hertz, A., Laporte, G. and Mittaz, M. *A Tabu Search Heuristic for the Capacitated Arc Routing Problem,* Operations Research. Vol. 48, No. 1, 2000, pp. 129-135.

[4] Hertz, A., and Mittaz, M. *A Variable Neighborhood Descent Algorithm for the Undirected Capacitated Arc Routing Problem,* Transportation Science. Vol. 35, No. 4, 2001, pp. 425-432.

[5] Longo, H., Poggi de Aragao, M. V. S. and Uchoa, E.: *Solving Capacitated Arc Routing Problems Using a Transformation to the CVRP.* Computers and Operations Research, Vol. 33, No. 6, 2006, pp. 1823-1837.

[6] http://www.uv.es/ belengue/carp.html

Solving Multicriteria Decision Making Problems using Microsoft Excel

Radomir Perzina[1], Jaroslav Ramik[2]

Abstract. The Analytic Hierarchy Process is frequently used approach for solving decision making problems. There exists wide range of software utilizing that approach. Main disadvantage of those software products is that they are commercial and relatively expensive and thus it prevents them to be used by small companies, students or researchers.
This paper introduces a Microsoft Excel add-in called DAME. Comparing to other programs DAME is free, can work with scenarios or multiple decision makers, allows for easy manipulation with data and utilizes capabilities of Microsoft Excel. Users can structure their decision models into three levels – scenarios/users, criteria and variants. There are provided three different methods for the evaluation of the weights of criteria, the variants as well as the scenarios/users – Saaty's Method, Geometric Mean Method and Fuller's Triangle Method. Multiplicative and additive syntheses are supported. All calculations are instant so users can easily see what happen if anything is modified. Bar chart is used for final ordering representation. The proposed software package is demonstrated on couple of illustrating examples of real life decision problems.

Keywords: analytic hierarchy process, multi-criteria decision making, pair-wise comparisons, Microsoft Excel.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Decision making in situations with multiple variants is an important area of research in decision theory and has been widely studied e.g. in [2], [3], [5], [7], [9], [10], [11], [12]. There exists wide range of computer programs that are able to help decision makers to make good decisions, e.g. Expert Choice (http://www.expertchoice.com), Decisions Lens (http://www.decisionlens.com), Mind Decider (http://www.minddecider.com), MakeItRational (http://makeitrational.com) or Super Decisions (http://www.superdecisions.com). Main disadvantage of those programs is that they are commercial and relatively quite expensive and thus it prevents them to be used by small companies or individual entrepreneurs.

Here we introduce a new Microsoft Excel add-in named DAME – Decision Analysis Module for Excel. Comparing to other software products for solving multicriteria decision problems, DAME is free, able to work with scenarios or multiple decision makers, allows for easy manipulation with data and utilizes capabilities of widespread spreadsheet Microsoft Excel. Users can structure their decision models into three levels – scenarios/users, criteria and variants. Standard pair-wise comparisons are used for evaluating both criteria and variants. For each pair-wise comparison matrix there is calculated an inconsistency index. There are provided three different methods for the evaluation of the weights of criteria, the variants as well as the scenarios/users - Saaty's Method [11], Geometric Mean Method [1] and Fuller's Triangle Method [2]. Multiplicative and additive syntheses are supported.

2 Software Description

DAME works with all current versions of Microsoft Excel from version 97. It consists of two individual files:

- DAME.xla – main module with user interface, it is written in VBA (Visual Basic for Applications),
- DAME.xll – it contains special user defined functions used by the application, it is written in C# and linked with Excel-DNA library (http://exceldna.codeplex.com).

[1] School of Business Administration, Silesian University, Department of Mathematical Methods in Economy, University Square 1934/3, Karviná, Czech Republic, perzina@opf.slu.cz
[2] School of Business Administration, Silesian University, Department of Mathematical Methods in Economy, University Square 1934/3, Karviná, Czech Republic, ramik@opf.slu.cz.

Both files must be placed in the same folder and macros must be permitted before running the module (see Microsoft Excel documentation for details). DAME itself can be executed by double clicking on the file DAME.xla. After executing the add-in there will appear a new menu item "DAME" in the Add-ins ribbon (in older Excel versions the menu item "DAME" will appear in the top level menu). A new decision problem can be generated by clicking on "*New problem*" item in the main DAME menu, see figure 1.

Figure 1 New problem menu

Then there will be shown a form with main problem characteristics, see figure 2.

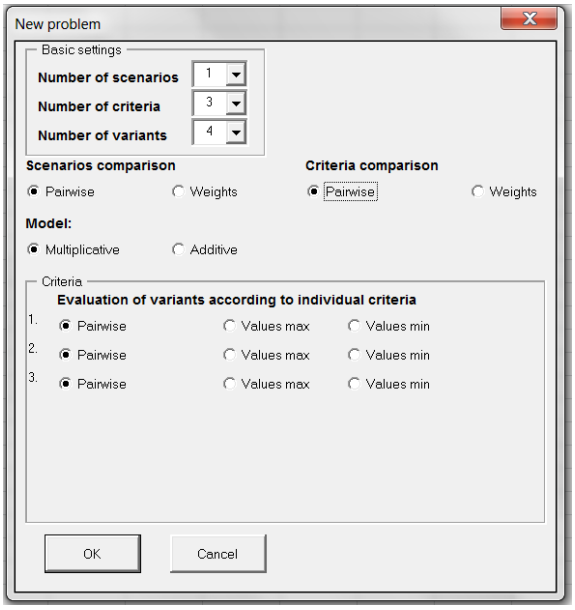


Figure 2 New problem characteristics

In the top panel there are basic settings: Number of scenarios, criteria and variants. In case a user doesn't want to use scenarios or there is just a single decision maker, the number of scenarios should be set to one. In the second panel we can set how we want to compare scenarios/users and criteria either using pairwise comparison matrix or set weights directly. Here we can also choose multiplicative or additive synthesis model. In the last panel users can chose how they want to evaluate variants according to individual criteria. There are three options: Pairwise – each pair of variants is compared individually, Values max – indicates maximization criterion where each variant is evaluated by single value, e.g. price and Values min – indicates minimization criterion where each variant is evaluated by single value, e.g. costs. When user confirms his options a new Excel sheet with forms is created, where user can set names of all elements and evaluate criteria and variants using pairwise comparison matrices as shown on figure 3.

Criteria	Crit 1	Crit 2	Crit 3	0.000	*Criteria weights*
Crit 1	1				0.333333
Crit 2	0	1			0.333333
Crit 3	0	0	1		0.333333

Figure 3 Pairwise comparison matrix

In the pairwise comparison matrix users enter values only in the upper triangle. The values in the lower triangle are reciprocal and automatically calculated. If criterion (variant) in the row is more important than the criterion (variant) in the column user enters values from 2 to 9 (the higher the value is the more important is the criterion in the row). If criterion (variant) in the row is less important than the criterion (variant) in the column user enters values from 1/2 to 1/9 (the less the value is the less important is the criterion in the row). If criterion (variant) in the row is equally important to the criterion (variant) in the column user enters value 1 or leaves it empty. In the top right corner there is calculated inconsistency index which should be less than 0.1, if it is greater we should revise our pairwise comparisons, so that they are more consistent. In the very right column there are

calculated weights of individual criteria (variants) based on the values in the pairwise comparison matrix and selected evaluation method. The weights w_k based on geometric mean method are calculated using the equation (1).

$$w_k = \frac{\left(\prod_{j=1}^{n} a_{kj}\right)^{1/n}}{\sum_{i=1}^{n}\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}},\ k = 1,2,\ldots,n \tag{1}$$

where w_k is weight of k-th criterion (variant), a_{ij} are values in the pairwise comparison matrix, and n is number of criteria (variants).

The inconsistency index is calculated using the formula (2).

$$GCI = \frac{2}{(n-1)(n-2)}\sum_{i<j}\log^2\left(a_{ij}\cdot\frac{w_j}{w_i}\right) \tag{2}$$

When we are entering values in individual pairwise comparison matrices all weights are being instantly recalculated, so we can see immediate impact of our each individual entry. Matrix and graph with total evaluation of variants is then shown at the bottom of the sheet. The resulting vector of weights of the variants **Z** is given by the formula (3).

$$\mathbf{Z} = \mathbf{W}_{32}\mathbf{W}_{21}, \tag{3}$$

where $\mathbf{W_{21}}$ is the $n\times1$ matrix (weighing vector of the criteria), i.e.

$$\mathbf{W}_{21} = \begin{bmatrix} w(C_1) \\ \vdots \\ w(C_n) \end{bmatrix}, \tag{4}$$

and $\mathbf{W_{32}}$ is the $m\times n$ matrix:

$$\mathbf{W}_{32} = \begin{bmatrix} w(C_1,V_1) & \cdots & w(C_n,V_1) \\ \vdots & \cdots & \vdots \\ w(C_1,V_m) & \cdots & w(C_n,V_m) \end{bmatrix}, \tag{5}$$

where $w(C_i)$ is weight of the criterion C_i, $w(V_r,C_i)$ is weight of variant V_r subject to the criterion C_i.

3 Case Study

Here we demonstrate the proposed add-in DAME on a decision making situation buying an “optimal” refrigerator with 3 decision criteria and 3 variants. The goal of this realistic decision situation is to find the best variant from 3 pre-selected ones according to 3 criteria: price (minimization criterion), efficiency (pairwise) and design (pairwise). At this stage we have just single decision maker, so the parameter Number of scenarios will be set to one. Setting of parameters can be seen on the figure 4.

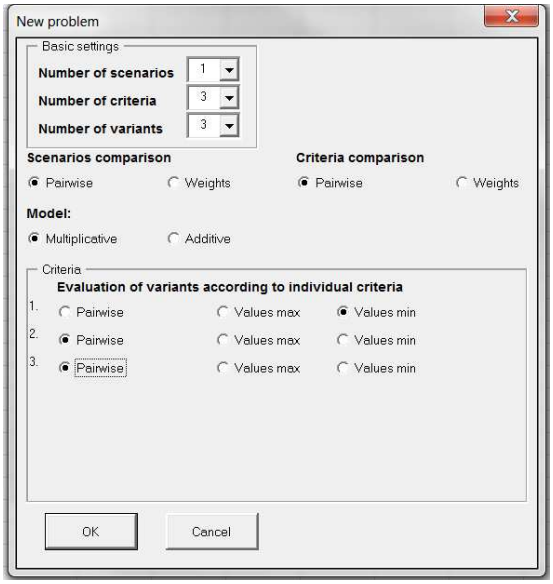


Figure 4 Case study – setting of parameters

When we submit the form a new sheet is generated. First we set names of criteria and variants, for simplicity we use default names for variants (Var 1, Var 2 and Var 3), see figure 5.

Names of criteria:		
Price	Efficiency	Design
Names of variants:		
Var 1	Var 2	Var 3

Figure 5 Case study – names of criteria and variants

Next step is comparison of individual criteria using pairwise comparison matrix with elements saying how much more important is criterion in the row than the criterion in the column, see figure 6.

Criteria	Price	Efficiency	Design	0.010	*Criteria weights*
Price	1	1	3		0.443429
Efficiency	1	1	2		0.387371
Design	0.333333	0.5	1		0.1692

Figure 6 Case study – criteria comparison

We can see that inconsistency index is less than 0.1 therefore we can say that our pairwise comparisons are consistent. In the very right column we can see calculated weights of individual criteria.

Final step is evaluation of variants according to individual criteria. Variants according the first criterion (price) will be evaluated by actual price and variants according the other two criteria (efficiency and design) will be evaluated using pairwise comparisons), see figure 7.

Price	Value				*Variants weights*		
Var 1	18				0.335329	0.229651	0.625013
Var 2	16				0.377246	0.12202	0.238487
Var 3	21				0.287425	0.648329	0.1365
Efficiency	Var 1	Var 2	Var 3	0.002			
Var 1	1	2	1/3				
Var 2	0.5	1	1/5				
Var 3	3	5	1				
Design	Var 1	Var 2	Var 3	0.010			
Var 1	1	3	4				
Var 2	0.333333	1	2				
Var 3	0.25	0.5	1				

Figure 7 Case study – evaluation of variants

As we can see both pairwise comparison matrices are consistent, because their inconsistency indexes are less than 0.1. In the top right matrix we can see calculated weights of all variants (rows) according to individual crite-

ria (columns). At this stage synthesis is calculated and we can see total evaluation of variants in the last table on figure 8 and graphical representation on figure 9. We can say that the best variant is Var 3 with weight 0.40 followed by Var 1 with weight 0.34 and the last one is Var 2 with weight 0.25.

CZn=	Weight	Rank
Var 1	0.343407	2
Var 2	0.254901	3
Var 3	0.401692	1

Figure 8 Case study – total evaluation of variants

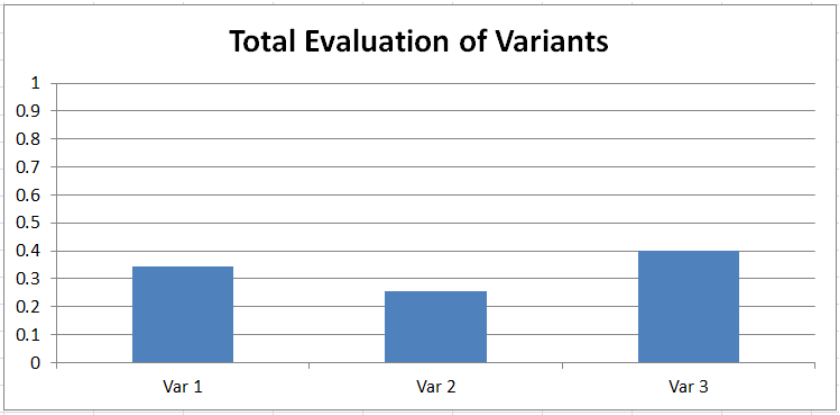


Figure 9 Case study – total evaluation of variants - graph

4 Case Study with Multiple Decision Makers

In real world people have different opinion on the same decision situation. Therefore our proposed software works also with multiple decision makers. In this case study we assume two decision makers – User A and User B. First we must compare both decision makers using pairwise comparison matrix. We believe that evaluations of user B are slightly "better" than evaluations of user A. The pairwise comparison matrix is shown on figure 10.

Scenarios	User A	User B	0.000	*Scenarios weights*
User A	1	1/2		0.333333
User B	2	1		0.666667

Figure 10 Case study – decision makers comparison

User A is using exactly the same entries as in the previous case study, so we just need to compare criteria and evaluate variants to individual criteria for the second user B, see figure 11.

Criteria Comparison:

Criteria	Crit 1	Crit 2	Crit 3	0.002	*Criteria weights*
Crit 1	1	3	5		0.648329
Crit 2	0.333333	1	2		0.229651
Crit 3	0.2	0.5	1		0.12202

Evaluation of Variants According to Individual Criteria:

Crit 1	Value	*Variants weights*		
Var 1	17	0.343348	0.332516	0.625013
Var 2	16	0.364807	0.139648	0.238487
Var 3	20	0.291845	0.527836	0.1365

Crit 2	Var 1	Var 2	Var 3	0.030
Var 1	1	3	1/2	
Var 2	0.333333	1	1/3	
Var 3	2	3	1	

Crit 3	Var 1	Var 2	Var 3	0.010
Var 1	1	3	4	
Var 2	0.333333	1	2	
Var 3	0.25	0.5	1	

Figure 11 Case study – criteria comparison and evaluation of variants – user B

Final evaluation of variants for user B can be seen on figure 12.

Zn=	Weight	Rank
Var 1	0.375229	1
Var 2	0.297685	3
Var 3	0.327086	2

Figure 12 Case study – final evaluation of variants – user B

Finally from both decision makers there is calculated synthesis and total evaluation of variants is shown on figure 13.

CZn=	Weight	Rank
Var 1	0.364622	1
Var 2	0.283424	3
Var 3	0.351955	2

Figure 13 Case study – total evaluation of variants

Comparing to the previous case study with just single decision maker we can see that final rank of variants has changed. Now the best variant is Var 1 with weight 0.36, then Var 3 with weight 0.35 and the last one Var 2 with weight 0.28.

5 Conclusions

In this paper we have proposed a new Microsoft Excel add-in DAME for solving decision making problems. Comparing to other decision support programs DAME is free, able to work with scenarios or multiple decision makers, allows for easy manipulation with data and utilizes capabilities of widespread spreadsheet Microsoft Excel. On two realistic case studies we have demonstrated its functionality in individual steps. This add-in is used by hundreds of students in the course Decision Analysis for Managers at the School of Business Administration in Karvina, Silesian University in Opava. It can be recommended also for other students, researchers or small companies.

Acknowledgements

This research was supported by the grant project of GACR No. 14-02424S.

References

[1] Aguaron, J., Moreno-Jimenez, J.M.: The geometric consistency index: Approximated thresholds. *European Journal of Operational Research* 147 (2003), 137-145.

[2] Fishburn, P. C.: A comparative analysis of group decision methods, *Behavioral Science* (16), 1971, 538-544.

[3] Gass, S.I., Rapcsák, T.: Singular value decomposition in AHP. *European Journal of Operational Research* 154 (2004), 573–584.

[4] Horn, R. A., Johnson, C. R.: *Matrix Analysis*, Cambridge University Press, 1990.

[5] Mazurek, J., Fiedor, J.: The Decision Tool for the Ordinal Consensus Ranking Problem. In: *Conference Proceedings of the International Scientific Conference ICT for Competitiveness 2012*. Silesian University, School of Business Administration, Karvina, 2012, 184-189.

[6] Ramík, J., Korviny, P.: Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean. *Fuzzy Sets and Systems* 161 (2010), 1604–1613.

[7] Ramik, J., Perzina, R.: Fuzzy ANP – a New Method and Case Study. In: *Proceedings of the 24th International Conference Mathematical Methods in Economics 2006*, University of Western Bohemia, 2006.

[8] Ramik, J., Perzina, R., DAME - Microsoft Excel Add-in for Solving Multicriteria Decision Problems with Scenarios. In: *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012*, Silesian University, School of Business Administration, 2012.

[9] Ramík, J., Vlach, M.: *Generalized concavity in optimization and decision making*. Kluwer Publ. Comp., Boston-Dordrecht-London, 2001, 305 p.

[10] Saaty, T.L.: Exploring the interface between hierarchies, multiple objectives and fuzzy sets. *Fuzzy Sets and Systems* 1, 1978, 57-68.

[11] Saaty, T.L.: *Multicriteria decision making - the Analytical Hierarchy Process*. Vol. I., RWS Publications, Pittsburgh, 1991.

[12] Saaty, T.L.: *Decision Making with Dependence and Feedback – The Analytic Network Process*. RWS Publications, Pittsburgh, 2001.

Matrix permutation problem for fair workload distribution

Štefan Peško[1], Roman Hajtmanek[2]

Abstract. We deal with solving an NP-hard problem which occurs in fair scheduling of workload distribution. Mathematical formulation of our problem is as follows: suppose we are given non-negative real matrix and positive weights of rows of that matrix. The problem is to find such permutations of the elements in columns of the matrix so that: (i) the irregularity measure of vector containing row sums of the permuted matrix is minimal, (ii) total cost of weighted row-sum of permuted matrix is minimal. We study an integer programming formulation of the problem. Computational experiments with heuristic, stochastic algorithm based on repeated solving of aggregated two-columns sub-problems are presented.

Keywords: fair schedule, uniform split, irregularity measure, matrix permutation problem

JEL classification: C02, C61, C65, C68, J33
AMS classification: 90B35, 11C99, 05D40

1 Introduction

We deal with solving an NP-hard problem which occurs in fair scheduling of workload distribution. It means that we need to split total workload to given number of approximately equal parts. The level of fairness is expressed by irregularity measure. The lower measure implies the higher regularity. The goal is to minimize the total cost of weighted parts of this fair schedule.

First formulations of Matrix Permutation Problem (MPP) can be found in Černý [3] (in Slovak) and Tegze and Vlach [8] (in English). They are motivated by Peško's mathematical interpretation of a real problem, when we needed to assign weekly work schedule to a given number of drivers as evenly as possible. The goal is to find a minimally varied set of row-sums of column permuted matrix. The approaches based on graph theory solving graph version of the MPP are studied by Czimmermann [1, 2]. For reference, 32 MPPs are defined and examined in work Dell'Olmoa [4] . They are characterized by particular measure of set uniformity to be optimized. 21 of the studied problems can be solved by linear time algorithms, 7 require more complex algorithms but can still be solved in polynomial time, and 3 are proved to be NP-hard.

The paper by Peško and Černý [6] presents different real situations when making managerial decision can be influenced by the methodology of fair assignment. The MPP with interval matrix presented in Peško [7] is motivated by practical need of regularity in job scheduling when inputs are given by real intervals. It is shown that this NP-hard problem is solvable in polynomial time for two-column interval matrix. Probability algorithm for general case of problem based on aggregated two-column sub-problem is presented. This algorithm for a real MPP is compared in Hajtmanek [5] to greedy heuristic.

Our model is motivated by following practical fair job scheduling problem:
Lets have m vehicles (garbage trucks) and n days. Every day a vehicle perform a work, so for each day we have m works (jobs) to be done. Suppose that elements of given matrix a_{ij} are benefits gained by performing work by i^{th} vehicle on j^{th} day. Total benefit of i^{th} vehicle is $s_i = \sum_{j=1}^{m} a_{ij}$. If there is a big difference between the gains $s_1, s_2, \ldots, s_m$, the drivers of vehicles will protest against it. Let w_i be unit cost of driver's work on i^{th} vehicle. Hence the employer first permutes elements in the individual

[1]University of Žilina, Univerzitn 8215/1, Žilina, Slovakia, Stefan.Pesko@fri.uniza.sk
[2]University of Žilina, Univerzitn 8215/1, Žilina, Slovakia, Roman.Hajtmanek@fri.uniza.sk

columns of matrix A in order to reach the row sums as regular as possible and second assign this schedule to drivers with minimum total cost.

We begin with studying mathematical formulation of the problem.

2 Mathematical formulation

The mathematical formulation of our problem is as follows: suppose we are given non-negative real matrix and positive weights of rows of matrix. The problem is to find permutations of the elements in columns of matrix so (i) the irregularity measure of vector containing row sums of the permuted matrix is minimal, (ii) total cost of weighted row-sum of permuted matrix is minimal.

More precisely we will deal with following optimization problem: For a real $m \times n$ matrix $\mathbb{A} = (a_{ij})$ and an ordered n-tuple $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ of permutations of the set $M = \{1, 2, \ldots, m\}$ let's $\mathbb{A}^\pi$ denote the real matrix whose element in the i^{th} row and j^{th} column is $a_{\pi_j(i),j}$ for $j \in N = \{1, 2, \ldots, n\}$. Let $\mathbb{S}^\pi = (s_1^\pi, s_2^\pi, \ldots, s_m^\pi)$ denote m-vector with elements

$$s_i^\pi = \sum_{j \in N} a_{\pi_j(i),j}, \quad i \in M,$$

whose i^{th} element is i^{th} row-sum of the permuted matrix $\mathbb{A}^\pi$.

Given a matrix $\mathbb{A} \in \Re^{m \times n}$ and real function $f : \Re^m \to \Re$ we consider following problem called *matrix permutation problem* (MPP):

$$f(\mathbb{S}^\pi_{\min}) = \min \Big\{ f(\mathbb{S}^\pi) : \pi \in \Pi_n, s_i^\pi = \sum_{j \in N} a_{\pi_j(i),j}, i \in M \Big\}, \tag{1}$$

where f is an irregularity measure and Π_n is the set of all n-tuples of permutations of the set M.

Following functions are examples of practical irregularity measures studied for the MPP in [6] (using denotation $x = x_1 + x_2 + \cdots + x_m$):

$$\begin{aligned}
f_{max}(x_1, x_2, \ldots, x_m) &= \max_{i \in M} x_i - \frac{x}{m}, \\
f_{min}(x_1, x_2, \ldots, x_m) &= \frac{x}{m} - \min_{i \in M} x_i, \\
f_{dif}(x_1, x_2, \ldots, x_m) &= \max_{i \in M} x_i - \min_{i \in M} x_i, \\
f_{sqr}(x_1, x_2, \ldots, x_m) &= \sum_{i \in M} x_i^2 - \frac{x^2}{m}.
\end{aligned}$$

Let's have optimum solution of the MPP in the form $\mathbb{S}^\pi_{\min} = (s_1^\pi, s_2^\pi, \ldots, s_m^\pi)$ and m-tuple of positive real numbers $\mathbb{W} = (w_1, w_2, \ldots, w_m)$ satisfying the conditions

$$w_1 \le w_2 \le \cdots \le w_m, \tag{2}$$

$$s_1^\pi \ge s_2^\pi \ge \cdots \ge s_m^\pi. \tag{3}$$

Then pair $(\mathbb{S}^\pi_{\min}, W)$ will be called optimum solution of the *fair matrix permutation problem* (FMPP). As we can see the ordering (2) and (3) minimise weighted row-sum of the MPP

$$g(\mathbb{S}^\pi_{\min}, W) = \sum_{i \in M} w_i s_i^\pi. \tag{4}$$

Note that the optimal solution of the FMPP in our interpretation of the job scheduling problem represents solution for drivers very well but it does not do this for solution with minimum cost for the employer. This minimum *unfair* cost is equal $g_u(\mathbb{S}^\pi, \mathbb{W}) = \sum_{i \in M} w_i s_i^\pi$ where every j^{th}-column of matrix $\mathbb{A}^\pi$ is a priori sorted by

$$a_{\pi_j(1),j} \ge a_{\pi_j(2),j} \ge \cdots \ge a_{\pi_j(m),j}.$$

In next text we will restrict to the *sqrt*-irregularity measure f_{sqr} which is more sensitive than measures f_{max}, f_{min} and f_{dif}.

3 Model of bivalent quadratic programming

We can see from formulation (1) that the main problem for solving the FMPP is exactly solving the MPP. Let value of bivalent variable be $x_{ijk} = 1$ if $\pi_j = k$ and $x_{ijk} = 0$ otherwise. This can be formulated and solved as following version of quadratic assignment problem (QAP):

$$\sum_{i\in M}\Big(\sum_{(j,k)\in N\times M} a_{ik}\cdot x_{ijk}\Big)^2 \to \min, \tag{5}$$

$$s.t. \sum_{k\in M} x_{ijk} = 1 \qquad \forall (i,j)\in M\times N, \tag{6}$$

$$\sum_{i\in M} x_{ijk} = 1 \qquad \forall (k,j)\in M\times N, \tag{7}$$

$$x_{ijk}\in\{0,1\} \qquad \forall (i,j,k)\in M\times N\times M. \tag{8}$$

The objective function (5) is the irregularity measure f_{sqr} only when the left side is constant. Constraints (6)–(8) define assignment polynom for every column of matrix $\mathbb{A}$.

Let's have optimal solution $\mathbb{X}^* = (x^*_{ijk})$ of the QAP. Then we can easily compute row-sum of permuted matrix $s^*_i = \sum_{j\in N}\sum_{k\in M} a_{ik}\cdot x^*_{ijk}$. After sorting it in descending order, given by permutation ψ of set M

$$s^*_{\psi(1)} \geq s^*_{\psi(2)} \geq \cdots \geq s^*_{\psi(m)}, \tag{9}$$

we got, in accordance with the definition of(4), the optimal solution of the FMPP.

So the optimal solution of the FMPP is possible to be found in two main steps: firstly the QAP is solved and secondly corresponding row-sum is resorted according to the ordering (9). Difficulty of the exact solving of the FMPP is that the corresponding QAP is NP-hard. However, there exists polynomial solvable instances on which is based following heuristic algorithm.

4 Algorithms

The first polynomial solvable instance of the MPP from 1986 is based on a theorem which was proved for more general assumptions on a class of irregularity measures. It can be used in general cases as a decomposition heuristic which repeatedly exact solves of aggregate two-columns instances.

4.1 Two-column case

For two-column case of the MPP, where $\mathbb{A}\in\Re^{m\times 2}$, we can find optimal solution by solving the following linear assignment problem (LAP):

$$\sum_{i\in M}\sum_{j\in M}(a_{i1}+a_{j2})^2\cdot y_{ij} \to \min \tag{10}$$

$$s.t. \sum_{j\in M} y_{ij} = 1 \qquad \forall i\in M, \tag{11}$$

$$\sum_{i\in M} y_{ij} = 1 \qquad \forall j\in M, \tag{12}$$

$$y_{ij}\in\{0,1\} \qquad \forall (i,j)\in M\times M. \tag{13}$$

We set $\pi_2(i) = j$ as value of variables $y_{ij} = 1$. We define $\pi_1(i) = i$ for $i\in M$ as permutation of first column. It is known that the linear assignment problem of size m is solvable in time $O(m^3)$ and so the two-column case is polynomial solvable. But our special cost of assign in object function (10) enables using known procedure *quicksort* with the time complexity $O(m\log m)$.

Theorem 1 (Tegze–Vlach [8])**.** *The matrix $\mathbb{A}\in\Re^{m\times 2}$ is an optimal permuted matrix of the MPP with the objective funcion f_{sqr} iff*

$$a_{11}\leq a_{21}\leq\cdots\leq a_{m1}, \tag{14}$$

$$a_{12}\geq a_{22}\geq\cdots\geq a_{m2}. \tag{15}$$

Example 1. Let's have 5 vehicles with drivers. For two working days we have given plain length of vehicle routing [km] by matrix $\mathbb{A}$ and we know matrix $\mathbb{W}$ of unit cost of drivers [€/km].

$$\mathbb{A} = \mathbb{A}^{\pi} = \begin{pmatrix} 85 & 95 \\ 80 & 65 \\ 70 & 50 \\ 45 & 40 \\ 20 & 35 \end{pmatrix} \quad \mathbb{W} = \begin{pmatrix} 10 \\ 12 \\ 12 \\ 13 \\ 15 \end{pmatrix} \quad \mathbb{S}^{\pi} = \begin{pmatrix} 180 \\ 145 \\ 120 \\ 85 \\ 55 \end{pmatrix}.$$

Columns of matrix $\mathbb{A}$ are sorted in descending order and so corresponding row-sum matrix $\mathbb{S}$ is sorted in this order too. Matrix $\mathbb{W}$ is sorted in ascending order and so for the employer is optimum assign i^{th} driver to i^{th} sum-jobs with cost with minimum but *unfair* total cost

$$g_u(\mathbb{S}^{\pi}, \mathbb{W}) = \sum_{i=1}^{5} w_i s_i^{\pi} = 10 \cdot 180 + 12 \cdot 145 + \cdots + 15 \cdot 55 = 6760 \text{ [€]}.$$

The drivers will protest because one driver work 180 km but one 55 km only. Optimum fair solution can be found by solving the MPP using theorem 1.

$$\mathbb{A} = \begin{pmatrix} 85 & 35 \\ 80 & 40 \\ 70 & 50 \\ 45 & 65 \\ 20 & 95 \end{pmatrix} \quad \mathbb{S}^{\pi} = \begin{pmatrix} 120 \\ 120 \\ 120 \\ 110 \\ 115 \end{pmatrix}.$$

Then we reorder matrix S in descending order and so we get fair solution of the FMPP

$$\mathbb{A}^{\pi} = \begin{pmatrix} 85 & 35 \\ 80 & 40 \\ 70 & 50 \\ 20 & 95 \\ 45 & 65 \end{pmatrix} \quad \mathbb{W} = \begin{pmatrix} 10 \\ 12 \\ 12 \\ 13 \\ 15 \end{pmatrix} \quad \mathbb{S}^{\pi}_{min} = \begin{pmatrix} 120 \\ 120 \\ 120 \\ 115 \\ 110 \end{pmatrix}.$$

with minimum *fair* cost

$$g(\mathbb{S}^{\pi}_{\min}, W) = \sum_{i=1}^{5} w_i s_i^{\pi} = 10 \cdot 120 + 12 \cdot 120 + \cdots + 15 \cdot 110 = 7225 \text{ [€]}.$$

4.2 General case

Good heuristic solution of the FMMP with the number of rows and columns greater than 2 can be found via following stochastic algorithm – decomposition heuristic:

Input: Matrices $\mathbb{A} \in \Re^{m \times n}$, $\mathbb{W} \in \Re^{m}$.

Output: Permuted matrix $\mathbb{A}^{\pi}$.

Step 1: Let $M = \{1, 2, \ldots, m\}, N = \{1, 2, \ldots, n\}$ and arbitrary $\pi = (\pi_1, \pi_2, \ldots, \pi_n) \in \Pi_n$ on set M.

Step 2: Choose randomly a non-empty subset $J \neq \emptyset$ of the set of columns $J \subset N$.

Step 3: Solve the two-column MPP with matrix $\mathbb{B} = (b_{ij})$ where

$$b_{i1} = \sum_{j \in J} a_{\pi_j(i),j}, \; b_{i2} = \sum_{j \in N-J} a_{\pi_j(i),j}$$

with optimal columns permutations ϕ_1, ϕ_2 that satisfied conditions (14) and(15)

$$b_{\phi_1(1)1} \leq b_{\phi_1(2)1} \leq \cdots \leq b_{\phi_1(m)1}, \; b_{\phi_2(1)2} \geq b_{\phi_2(2)2} \geq \cdots \leq b_{\phi_2(m)2}.$$

Step 4: Update n-tuple π apply permutation ϕ_1 resp. ϕ_2 to π_j for $j \in J$ resp. $j \in M - J$.

Step 5: If no stop criterion is satisfied GOTO Step 2.

Step 6: Update matrix $\mathbb{A}^\pi$ by resorting its row-sums in descending order (9) and STOP($\mathbb{S}^\pi_{\min}, W$).

The stop criterion in step 5 may be the maximal number of iterations (steps 2–5), the given number of consecutive unsuccessful iterations, maximum runtime etc. Steps 1–5 solve heuristically the MPP only in general. Last step 6 is necessary for solution $(\mathbb{S}^\pi_{\min}, W)$ of the FMPP considering after this step satisfy conditions (2) and (3). In case when the MPP is solved exactly, the FMPP is solved exactly too.

5 Computation experiments

The computations reported in this section have been carried out on a 2-core I5, 2.5 Ghz, RAM 2 GB. The first exact solution was done in Gnumeric [1] spreadsheet for small instances of randomly generated matrix. This spreadsheet has only solver for linear model (LP/MILP) at so we have to reformulate the 3-columns QAP problem with the matrix $\mathbb{A} \in \Re^{m\times 3}$ as following covering problem (CP3):

$$\sum_{(i,j,k,l)\in M^4} (a_{ij} + a_{ik} + a_{il})^2 \cdot z_{ijkl} \to \min, \tag{16}$$

$$s.t. \sum_{(j,k,l)\in M^3} z_{ijkl} = 1 \qquad \forall i \in M, \tag{17}$$

$$\sum_{(i,k,l)\in M^3} z_{ijkl} = 1 \qquad \forall j \in M, \tag{18}$$

$$\sum_{(i,j,l)\in M^3} z_{ijkl} = 1 \qquad \forall k \in M, \tag{19}$$

$$\sum_{(i,j,k)\in M^3} z_{ijkl} = 1 \qquad \forall l \in M, \tag{20}$$

$$z_{ijkl} \in \{0,1\} \qquad \forall (i,j,k,l) \in M^4. \tag{21}$$

The value of variables $z_{ijkl} = 1$ defines three mappings of permutations $\pi_1(i) = j, \pi_2(i) = k, \pi_3(i) = l$. Row-sum of elements in i^{th} row of permuted matrix $\mathbb{A}^\pi$ is given $s_i = \sum_{(j,k,l)\in M^3}(a_{ij} + a_{ik} + a_{il})z_{ijkl}$.

m	5	10	15	20	25	30	35	40
CP3	0 : 41	0 : 57	1 : 12	3 : 01	13 : 38	41 : 10	112 : 21	130 : 09
FMMP	0 : 46	1 : 07	1 : 20	3 : 10	13 : 47	41 : 21	112 : 33	130 : 21

Table 1 *Mean runtime [MIN:SEC] of solving the FMMP via CP model for $m \times 3$ instances in Gnumeric.*

Results of computation experiments in table 1 with 20 randomly generated instances for every type of matrices $m \times 3$ show that with growing number of rows mean runtime grows exponentially. Absolute runtime difference of the last step from CP solution to the FMPP solution is about 10 seconds only.

We used Python[2] API for implementation of our decomposition heuristic. For experimental estimation of the quality of the solution we first generated larger instances with known optimal solution. Instances were generated so that part $m \times (n-1)$ of submatrix is generated randomly with normal distribution with the same mean and variance in columns. Last column of matrix $m \times n$ is evaluated as a maximum row-sum of $m \times (n-1)$ submatrix minus its row-sum. After this operation every row-sum of matrix A has the same value. Then we randomly permute the elements in columns and subsequently sort the column of matrix under decreasing order. After this operations it is very hard to find for matrix $\mathbb{A}$ permuted matrix $\mathbb{A}^\pi$ with the same row-sums.

Experiments in table 2 show runtime of solving the FMPP for 10 the above generated instances for every type of matrix. That symbol $\bullet$ notes case when heuristic at least one from ten instances do not find optimum solution. We use $100mn$ of consecutive unsuccessful iterations as a stop criterion of heuristic.

[1] http://www.gnumeric.org
[2] https://www.python.org

$m \times n$	5	10	15	20	25	30	35	40
10	0 : 26	0 : 37	0 : 42	1 : 10*	1 : 47	1 : 31	1 : 39•	1 : 50
20	0 : 41	0 : 23	0 : 57	1 : 42	2 : 36	2 : 51	2 : 09	2 : 18
30	1 : 06	0 : 55	1 : 20	2 : 11	2 : 43	3 : 11•	3 : 28•	4 : 14
40	2 : 28	0 : 57•	1 : 42	3 : 10	2 : 47•	3 : 31	3 : 30	3 : 58

Table 2 *Mean runtime [MIN:SEC] for the FMMP via decomposition heuristic with $m \times n$ instances.*

Open question is: Why are presented results so good? Note that when we extended maximal runtime to 45 minutes all computed examples were executed to optimum.

6 Conclusion

The paper has shown that the new original formulation of the matrix permutation problem for fair workload distribution can be correctly solved by stochastic algorithm based on repeated solving aggregated two-columns subproblems.

Acknowledgements

The research was supported by the research grants VEGA 1/0296/12 Public Service Systems with Fair Access to Service and APVV-0760-11 Designing of Fair Service Systems on Transportation Networks.

References

[1] Czimmermann, P.: The Most Regular Perfect Matchings in Graph, *Studies of the University of Žilina*, Mathematical Series, Vol. **16** no. 1, (2003), pp. 15-18.

[2] Czimmermann, P.: A Note on Using Graphs in Regular Scheduling Problems, *Communications*, Scientific letters of the University of Žilina, no. **4**, (2003), pp. 47-48.

[3] Černý, J.: How to minimise non-uniformity, (in Slovak), *PMFA XXX*, (1985), pp. 145-150.

[4] Dell'Olmoa, P., Hansenb, P., Pallottinoc, S., Storchid G.: On Uniform k-Partition Problems, *Discrete Applied Mathematics*, Vol. **150**, Issues 13, (2005), pp. 121-139.

[5] Hajtmanek, R.: Heuristic algorithms for uniform schedules of matrix, (in Slovak), In: *Otvorený softvér vo vzdelávaní, výskume a v IT riešeniach*, OSSConf2013, Žilina, (2013), pp. 33-37.

[6] Peško, Š., Černý, J.: Uniform Splitting in Managerial Decision Making, *E+M, Economics and Management* IX, vol.**4**, (2006), Technical university of Liberec, pp. 67-71.

[7] Peško, Š.: The Matrix Permutation Problem in Interval Arithmetic. *Journal of Information, Control and Management Systems*, Vol. **4**, (2006), pp. 35-40.

[8] Tegze, M., Vlach, M.: On the matrix permutation problem, *Zeitschrift fr Operations Research*, Vol. **30**, Issue 3, (1986), pp. A155-A159.

Evaluation of mutual fund performance on Polish capital market with the use of market timing models

Radosław Pietrzyk[1]

Abstract. Performance analysis of mutual funds is a very important theoretical and practical issue. There is no consensus on the market, which methods of investment performance should be used. This study examines performance of Polish mutual funds between 2001 and 2013. The mutual fund data set contains returns of Polish equity funds having at least 80% of the fund capital invested in domestic equity. Regression based market timing models (e.g.: Treynor-Mazuy (1966), Henriksson-Merton (1981) and Connor-Korajczyk (1991)) are used to assess market timing and stock selection abilities of fund managers. The main purpose of the article is to compare different methods of evaluating fund abilities in the area of market timing and security selection.
Estimated parameters of the models used in the research are statistically significant and the models are well fitted to data. However, it is often impossible to find evidence of any market timing ability in the selected group of funds. The results show that Polish mutual funds exhibit selectivity ability more often than timing ability. Most of them have negative timing. The results of the study may be important both for investors and fund managers.

Keywords: Mutual fund performance, market timing, selectivity.

JEL Classification: C20**,** G11, G14, G23
AMS Classification: 62J02, 62J05, 91G10, 91G70

1 Introduction

The main purpose of the article is to examine abilities of the mutual fund managers in Poland. The performance of mutual funds is assessed with measures of selectivity and market timing abilities. These two groups of measures allow to assess if portfolio managers are able to forecast changes in market trends, suit their investment strategies to the changes and apply proper portfolio selection techniques so that they outperform other participants of the market. This gives a tool for comparison of actively managed portfolios with passive ones and with benchmarks of the funds. Using market timing models allows to verify a hypothesis that a mutual fund obtains on average returns that are above the average for the market indices. Comparing returns yielded by mutual funds with returns on stock market indices implies that only funds that allocate vast part of their assets in stock market are taken here into consideration. For the needs of this research an 80% minimum threshold of stock investments is assumed. Eleven mutual funds operating in Poland in the years 2001-2013 are taken into analysis.

2 Market timing models

Evaluation of the ability to use market trend forecasts, as well as to pick up stocks, is based on a linear single-index model, in which portfolio return in period t is influenced by an aggregated risk factor. In one of the simplest and also the most popular models of this type, proposed by Sharpe, the relationship is given with the following formula:

$$R_{pt} = \beta_p R_{mt} + e_{pt} \tag{1}$$

where
R_{pt} – portfolio rate of return in period t,
R_{mt} – market portfolio rate of return in period t,
β_p – measure of sensitivity of portfolio return to aggregated market risk factor (in this model manifested by changes in return on market portfolio),
e_{pt} – random residual in period t.

In this model market portfolio is treated as a natural benchmark for stock portfolio managers. One of its application is analysis of deviations of realized rates of returns from theoretical returns obtained from the model. If the difference between realized and theoretical returns is on average positive, it may be treated as an excess rate

[1] Wroclaw University of Economics, Department of Financial Investments and Risk Management, Komandorska 118/120, 53-345 Wroclaw, Poland, radoslaw.pietrzyk@ue.wroc.pl

of return resulting from over-average skills of portfolio managers. This was the concept underlying the Jensen's [6] α measure:

$$R_{pt} - R_{ft} = \alpha_p + \beta_p \left(R_{mt} - R_{ft} \right) + e_{pt} \tag{2}$$

where:
R_{ft} – risk-free rate of return in period t,
α_p – selectivity measure.

This measure may be interpreted as the difference between realized rate of return of the managed portfolio and the rate of return on a passive portfolio with the same risk.

The model presented above does not, however, allow to identify the source of the excess return and attributing it to particular skills of a portfolio manager. Jensen's α gives only synthetic information about these skills. On the basis of CAPM model there originated also models that are counted among the so called market timing models. The measures based on them may be divided into security-selection and market timing measures. The basis of these models are also regression equations, but, unlike Sharpe model, underlying CAPM, in these models it is assumed that portfolio composition may change in time. This generalization gave an opportunity to isolate the pieces of information that are needed to assess particular market-timing skills of portfolio managers. This includes, among others, not only assessment how good trend-predicting skills the managers show, but also how elastic their strategy is in reacting to changing situation.

One of the first propositions of this type was a model by Treynor and Mazuy [10]. This is a quadratic regression model. It is based on the assumption that, if the managers are able to predict market trend reversals, they will be also adjust portfolio composition to these changes. This should lead to overweighting risky assets (market portfolio) in the portfolio when price increases are expected and underweighting them if decreases in prices are expected. Treynor and Mazury proposed then to augment the model by a squared component, which made the relationship between the rate of return on the analyzed and market portfolio nonlinear:

$$R_{pt} - R_{ft} = \alpha_p + \beta_p \left(R_{mt} - R_{ft} \right) + \gamma_p \left(R_{mt} - R_{ft} \right)^2 + e_{pt} \tag{3}$$

where: γ_p – market timing coefficient.

A positive value of this coefficient means that the portfolio gives an extra return thanks to adjusting portfolio composition to changes of general trends in the stock market.

Another solution presented Henriksson and Merton [4] and Merton [7], whose model is based on two regressions. One of them corresponds to periods of lower trend (negative excess returns) and the second is for periods of upper trend (positive excess returns). Model constructed like this may be used for evaluating the manager's skill to adjust portfolio composition to short-term trends in the market. The ability of adjustment consists in increasing the risk exposure (higher beta) in periods of growing prices and reducing risk (lower beta) in periods of falling prices. This model may be also expressed in a compact version using one equation:

$$R_{pt} - R_{ft} = \alpha_p + \beta_p \left(R_{mt} - R_{ft} \right) + \gamma_P \max\left(0, \left(R_{mt} - R_{ft} \right)\right) + e_{pt} . \tag{4}$$

The value of $\beta - \gamma$ expression may be interpreted as beta coefficient analogue for the case of a downturn in the market. A positive value of γ parameter means that the manager is able to predict market trends and adjust portfolio composition to these predictions.

A modification of this approach was presented by Connor and Korajczyk [3] who proposed to generalize it by taking into account a cash flow from buying or selling a put option. Unlike the model described by the eq. (4), asymmetric instruments used here are non-costless. The model may be directly derived from the eq. (4) by substitution:

$$\alpha_P = \alpha_P^* - \left(1 + R_{ft}\right) \gamma_P P_0 \tag{5}$$

where:
P_0 – value of a put option on the market portfolio at the beginning of the investment period,
α_P^* – selectivity measure in Connor-Korajczyk model.

After substituting the eq. (5) to the eq. (4) the following formula is obtained:

$$R_{pt} - R_{ft} = \alpha_p^* + \beta_p \left(R_{mt} - R_{ft} \right) + \gamma_p nput + e_{pt} , \tag{6}$$

where:

$$nput = \max\left(0, R_{mt} - R_{ft}\right) - \left(1 + R_{ft}\right) P_0 . \tag{7}$$

The *nput* instrument may be described as a contract that hedges the portfolio and, at the same time, is responsible for asymmetry of the model. An investor that holds the portfolio and the *nput* instrument has a guarantee

that the payoff from this strategy will be equal to the risk free rate minus the (risk-free-rate-capitalized) future value of the put option. Research on the mutual fund market in Poland with the use of this model is presented, amongst others, by Pietrzyk [9].

A two-factor extension of these models was presented, amongst others, by Weigel [11], who, building on Henriksson-Merton model, took into consideration three classes of assets: stocks, bonds and risk free investments. The model is augmented by an additional bond market index. This approach seems particularly justified when mixed funds, investing in different classes of assets, are analyzed. The model is defined by the following equation:

$$R_{pt} - R_{ft} = \alpha_p + \beta_{sp}\left(R_{mt} - R_{ft}\right) + \beta_{bp}\left(R_{bt} - R_{ft}\right) + \gamma_p \max\left(0, R_{mt} - R_{ft}, R_{bt} - R_{ft}\right) + e_{pt} \ , \tag{8}$$

where:
γ – market timing coefficient,
R_{bt} – return on treasury bond market index in a period t,
R_{mt} – return on stock market index in the period t,
β_{bp} – beta coefficient of the portfolio with respect to the bond index,
β_{sp} – beta coefficient of the portfolio with respect to stock index.

Yet another solution was proposed, amongst others, by Jensen [5] and Bhattacharya and Pfleiderer [1], who took into account expected rates of return in the market. Coggin, Fabozzi and Rahman [2] pointed out that when estimating regression parameters, both in Treynor-Mazuy and Bhattacharya-Pfleiderer models, a heteroscedasticity correction should be made. They postulated that random residuals will show the property of conditional heteroscedasticity because of the attempts by portfolio managers to adjust investment strategy to market trend changes. Thus, this conditional heteroscedasticity should be observed even if subsequent returns on stocks were independent, identically distributed random variables.

3 Data and methodology

Analysis of portfolio management performance in the area of market timing and security picking was performed for some chosen mutual funds operating in Polish market since at latest 1st Jan. 2001 until at least the end of 2013. This choice allows to do the research for a long period. Taking a long sample helps to marginalize the influence of outliers on the results of evaluation. The funds selected for the analysis allocate most of their means in risky securities, mainly stocks. It is assumed that the funds taken into consideration should use benchmarks that contain one of Polish stock market indices in the proportion of at least 80%. The selected funds use benchmarks comprising two main indices of the Warsaw Stock Exchange, that is: Warsaw Stock Exchange Index (WIG) and Warsaw Stock Exchange Large-Cap Index of 20 major and most liquid companies in the WSE Main List (WIG20). The remaining part of the benchmark is interest rate of deposits in the interbank market. In the majority of mutual funds the benchmarks are moreover corrected by an asset management fee (expressed in percentage terms).

Table 1 Rates of returns of mutual funds and equity index WIG

Fund	*R* 2001-2013	*R* monthly	st. dev. monthly	*R* weekly	St. dev. weekly
ARKA	91.10%	0.55%	6.81%	0.13%	2.78%
BPH	85.67%	0.51%	6.07%	0.13%	2.54%
ING	74.06%	0.45%	6.39%	0.11%	2.68%
LEGG MASON	113.51%	0.68%	6.17%	0.17%	2.48%
NOVO	51.09%	0.32%	6.51%	0.07%	2.74%
PKO	54.09%	0.32%	6.56%	0.08%	2.57%
PZU	69.73%	0.42%	6.10%	0.10%	2.48%
SKARBIEC	96.22%	0.58%	6.18%	0.14%	2.56%
PIONEER	11.30%	0.07%	7.46%	0.02%	2.95%
UNI KORONA	110.34%	0.67%	6.35%	0.16%	2.72%
WIG index	105.55%	0.63%	7.06%	0.16%	2.91%
WIBOR	81,13%	0,48%			

The analysis spans the period of Jan. 1 2001 – Dec. 31 2013. Samples of weekly (5-day), as well as monthly (20-day), logarithmic rates of returns were used. As the risk-free rate, it was taken the annual offered rate of

Polish interbank market (WIBOR), adjusted to weekly and monthly capitalization periods. Two of the models discussed above, Henriksson-Merton and Treynor-Mazuy, are used in the research. There are calculations for Connor-Korajczyk model performer as well, but the results were almost identical with those of the Henriksson-Merton model, and thus the output was treated as redundant. The broadest index of Polish stock market, WIG, is taken as a benchmark.

During the analyzed period all the funds obtained a positive return, but only 2 of them had a return higher than that of the benchmark (WIG index). **Table 1** presents rates of returns of mutual funds in the whole period, as well as weekly and monthly mean returns and their risk measured with standard deviation.

It may be observed that 5 of the 10 mutual funds realized in this period lower weekly and monthly average returns than the risk free rate. From theoretical and practical point of view the results for monthly horizons are more important, because adjusting of investment strategy takes usually some time. This is why the results of the research are presented for monthly rates of return.

4 Empirical results

In the research there was used Treynor-Mazuy model and Henriksson-Merton model. As it has already been mentioned, calculations for Connor-Korajczyk model were also performed, but the output will not be further discussed because of its unanimity with Henriksson-Merton model results.

Table 2 Estimates for Treynor-Mazuy model – monthly rate of returns

Fund	Parameter	Estimation of parameter	Standard error	*t*-value	*p*-value	R^2	F statistic	Degr. of Freedom
Arka	γ	-0.217	0.120	-1.810	0.072	0.910	843.4	166
	β	0.898	0.025	35.813	0.000			
	α	0.000	0.002	0.247	0.805			
BPH	γ	-0.115	0.090	-1.272	0.205	0.936	1 215.1	166
	β	0.820	0.019	43.421	0.000			
	α	-0.000	0.001	-0.264	0.792			
ING	γ	0.108	0.073	1.472	0.143	0.962	2 100.5	166
	β	0.898	0.015	58.510	0.000			
	α	-0.002	0.001	-2.122	0.035			
LEGG MASON	γ	-0.200	0.070	-2.853	0.005	0.963	2 156.9	166
	β	0.839	0.015	57.291	0.000			
	α	0.002	0.001	1.702	0.091			
NOVO	γ	-0.444	0.086	-5.142	0.000	0.949	1 554.4	166
	β	0.852	0.018	47.244	0.000			
	α	-0.001	0.001	-0.551	0.583			
PKO	γ	-0.676	0.082	-8.248	0.000	0.955	1 759.4	166
	β	0.834	0.017	48.723	0.000			
	α	0.001	0.001	0.438	0.662			
PZU	γ	-0.233	0.085	-2.743	0.007	0.944	1 394.8	166
	β	0.815	0.018	45.849	0.000			
	α	-0.001	0.001	-0.556	0.579			
SKARBIEC	γ	-0.066	0.082	-0.805	0.422	0.949	1 560.8	166
	β	0.846	0.017	49.509	0.000			
	α	0.000	0.001	0.062	0.951			
PIONEER	γ	-0.298	0.086	-3.461	0.001	0.961	2 066.7	166
	β	1.005	0.018	55.751	0.000			
	α	-0.004	0.001	-3.369	0.001			
UNI KORONA	γ	0.063	0.081	0.778	0.438	0.953	1 686.9	166
	β	0,883	0,017	52,198	0,000			
	α	0,000	0,001	0,183	0,855			

Estimates of model parameters (**Table 2** and **Table 3** contain results for monthly returns) show that the models are well fitted to the data. Coefficients of determination R^2 are above 0.9 for monthly data and above 0.8 for weekly ones. *F*-statistics also indicate a satisfactory significance of the models. The analysis of statistical properties for particular coefficients in the models show that β parameter estimates are significant in all cases (both for Treynor-Mazuy and Henriksson-Merton model). Values of these parameters indicates a strong connection of the portfolios with the market index. Beta coefficients for monthly are at the level of 0.725-0.902 for Henriksson-Merton model and 0.82-1.005 for Treynor-Mazuy.

In 5 cases for the T-M model and in 6 cases for the H-M model there were no premises demanding rejection of the hypothesis that γ coefficient equals zero. Values of γ coefficient estimator show that it is hard to prove any additional skills of portfolio managers in the area of market timing based on general market trend prediction. In Treynor-Mazuy model, for monthly data, the γ coefficient was negative in 8 cases and in the two cases with apparently positive values of γ the estimates were statistically insignificant (*p*-values of 0,14 and 0,44). For weekly data the values of γ coefficient are negative in all cases. There are no such mutual funds that gammas are positive and significantly different from zero for both models. It means that there are no evidence that they try to time the market. Nor are there any premises supporting the hypothesis that they have abilities to do this.

Table 3 Estimates for Henriksson-Merton model – monthly rate of returns

Fund	Parameter	Estimation of parameter	Standard error	*t*-value	*p*-value	R^2	F statistic	Degr. of Freedom
Arka	γ	-0.038	0.069	-0.553	0.581	0.909	827.2	166
	β	0.896	0.047	19.025	0.000			
	α	0.000	0.002	0.116	0.907			
BPH	γ	-0.080	0.051	-1.575	0.117	0.936	1 221.8	166
	β	0.783	0.035	22.321	0.000			
	α	0.001	0.002	0.643	0.521			
ING	γ	0.023	0.042	0.560	0.576	0.962	2 076.5	166
	β	0.902	0.029	31.406	0.000			
	α	-0.002	0.001	-1.552	0.123			
LEGG MASON	γ	-0.099	0.040	-2.474	0.014	0.963	2 130.9	166
	β	0.798	0.027	29.125	0.000			
	α	0.003	0.001	2.310	0.022			
NOVO	γ	-0.174	0.051	-3.422	0.001	0.945	1 429.1	166
	β	0.789	0.035	22.592	0.000			
	α	0.002	0.002	0.861	0.390			
PKO	γ	-0.286	0.050	-5.684	0.000	0.947	1 478.2	166
	β	0.725	0.035	20.954	0.000			
	α	0.004	0.002	2.516	0.013			
PZU	γ	-0.106	0.049	-2.179	0.031	0.943	1 371.1	166
	β	0.773	0.033	23.188	0.000			
	α	0.001	0.002	0.506	0.614			
SKARBIEC	γ	-0.010	0.046	-0.206	0.837	0.949	1 554.8	166
	β	0.847	0.032	26.571	0.000			
	α	-0.000	0.002	-0.013	0.990			
PIONEER	γ	-0.130	0.050	-2.626	0.009	0.960	2 005.4	166
	β	0.954	0.034	28.051	0.000			
	α	-0.002	0.002	-1.338	0.183			
UNI KORONA	γ	0.017	0.046	0.365	0.716	0.953	1 681.9	166
	β	0.888	0.032	28.143	0.000			
	α	0.000	0.002	0.065	0.949			

For selectivity the estimators of the respective parameter are somewhat different than for market timing. The parameter is statistically significant only in two cases, both for the H-M and T-M model. For monthly data, there are 8 cases for Henriksson-Merton model in which hypothesis that α parameter is equal to zero cannot be reject-

ed and in two cases the null hypothesis is rejected (where *p*-values are of 0,02 and 0,01). In the cases in which the parameter is significant its value is positive. It may be, thus, concluded that these funds (Legg Mason, PKO) are able to generate excess returns thanks to security selection skills. For weekly data such a conclusion might be drawn in 4 cases (Legg Mason, PKO, PZU, Arka) (*p*-values from 0,0005 to 0,027).

For Treynor-Mazuy model and for monthly data there is only one case in which the value α parameter is higher than zero and statistically significant (Legg Mason – *p*-value=0,09). For weekly data there are two such cases (Arka, Legg Mason). For monthly data there are 2 funds with α parameter lower than 0 and statistically significant (ING, Pioneer). For weekly data there is only one such case (Pioneer).

5 Conclusions

Mutual fund portfolio management evaluation is an important subject also from practical point of view. The results obtained in this research support the hypothesis that portfolio managers in mutual funds do not achieve on average excess returns from market timing and security picking activity, as compared to returns from a passive strategy (index investing). Similar conclusions bring the research based on some other market timing models (comp. [8]). Performance analysis on Polish market was done, amongst others, in [12]. The results are however hardly comparable due to different time range.

Also the analysis of selectivity indicator does not detect much symptoms of over-average advantages from picking up securities based on some particular pricing or forecasting skills. Some rare cases of significant positive α show that only few portfolio managers are able to successfully make active bets when selecting assets to the portfolios.

References

[1] Bhattacharya S.,Pfleiderer P.: A note on performance evaluation, Technical Report 714, Stanford University, 1983.
[2] Coggin T.D., Fabozzi F.J., Rahman S.: The investment performance of U.S. equity pension fund managers: an empirical investigation, *The Journal of Finance*, Vol. 48, No 3 (1993), 1039-1055.
[3] Connor G., Korajczyk R.A.: The Attributes, Behavior and Performance of U.S. Mutual Funds. *Review of Quantitative Finance and Accounting*, Vol. 1, No. 1 (1991), 5-26.
[4] Henriksson R.D., Merton R.C.: On Market Timing And Investment Performance. II. Statistical Procedures For Evaluating Forecasting Skills. *Journal of Business*, Vol. 54, No. 4 (Oct., 1981), 513-533.
[5] Jensen M.C.: Optimal Utilization of Market Forecasts and the Evaluation of Investment Performance, In: *Mathematical Methods in Investment and Finance*, ed. Szego G.P., Shell K., North-Holland, Amsterdam 1972.
[6] Jensen M.C.: The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, Vol. 23, No 2 (1968), 389-416.
[7] Merton R.C.: On Market Timing And Investment Performance. I. An Equilibrium Theory Of Value For Market Forecasts. *Journal of Business*, Vol. 54, No. 3 (Jul., 1981), 363-406.
[8] Pietrzyk R.: Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, No.* 242 (2012), 291-305.
[9] Pietrzyk, R.: Efektywność inwestycji funduszy inwestycyjnych – wykorzystanie modeli „market timing". *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, No.176 (2011), 531-540.
[10] Treynor J.L., Mazuy K.: Can Mutual Funds Outguess the Market? *Harvard Business Review*, Vol. 44, No. 4 (1966), 131-136.
[11] Weigel E.J., The Performance of Tactical Asset Allocation, *Financial Analysts Journal*, Vol. 47, No. 5, (1991), 63-70
[12] Zamojska A., Timing - metody pomiaru i empiryczna weryfikacja na przykładzie polskich funduszy inwestycyjnych. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, No. 60 (2009), 525-532.

Optimization of financing multiple goals with multiple investment programs in financial planning for households

Radoslaw Pietrzyk [1], Pawel Rokita [2]

Abstract. Every household during its life cycle realizes several financial goals. The most important are usually the following five: retirement, buying a house, bringing up children, funding their education and leaving a bequest. These goals are characterized by different realization terms and magnitudes. Most households would not be able to afford financing them without previous saving and investing some part of their income (prefinancing). There are also such goals that just cannot be financed in any other way than prefinancing (e.g., retirement – no possibility of postfinancing with a credit). There arises a need of finding an optimal set of investment programs along with information which program is destined to cover which financial goal. In the article a method of facilitating life-long financial planning for a household is proposed. It allows to find the optimal match between systematic investment plans (SIP) with the same level of risk and multiple financial goals. The choice of portfolios of systematic investment plans is multi-criteria optimization with a number of constraints. At least for expected values of all stochastic factors, household financial liquidity should be preserved in all periods. The plan should also allow for accomplishment of all financial goals set by the household members. The main criterion is minimization of cost for each possible combination of goals and available investment programs. But the resulting life-long cumulated surplus trajectory should also suit other preferences of the household, like their life-length risk aversion. This is why for sets of financial programs chosen in one step of the optimization procedure, a goal function based on expected discounted utilities is maximized in one of the next steps. The result is a set of systematic investment plans with accompanying information which programs are destined to cover which financial goal.

Keywords: Multiobjective optimization, personal finance, asset selection, intertemporal choice, systematic investment plan.

JEL Classification: C61, G11, D14
AMS Classification: 91B16, 91G10

1 Introduction

When constructing a financial plan with a number of goals and multiple investment products available, two tasks are to be discussed. The first is selecting a combination from amongst available systematic-investment programs and finding a match between investment programs and the goals. It is allowed here that one program is used to finance one or more goals. The second task is optimization of term structure of household cash flows. It is assumed that the bunch of goals themselves is just declared by the household and not subject to optimization. Both tasks are strictly connected. Investment program mixes chosen in preliminary selection are evaluated then by putting them into the model of household cash flows and calculating goal function value for each. The goal function used in this research will be further referred to as *value function*, because it is a kind of generalization of utility function. The choice of investment program mix for which the value function gives the highest output involves consumption optimization. It consists in finding the optimal combination of two decision variables: (1) the consumption – investment proportion and (2) shares of particular household members in joint household investments. The optimization is performed under a number of constraints. Besides the main, budget constraint, to the constraints belong also the investment program mix (the one that is considered in a given run of the optimization procedure) and household goals to be financed by the mix (which are the same each time). The final selection of investment program mix is, thus, the result of using the optimization procedure many times for different pre-selected program mixes, where the mixes themselves are not decision variables, but rather parameters determining a part of constraints.

[1] Wroclaw University of Economics, Department of Financial Investments and Risk Management, ul. Komandorska 118/120, 53-345 Wroclaw, Poland, radoslaw.pietrzyk@ue.wroc.pl
[2] Wroclaw University of Economics, Department of Financial Investments and Risk Management, ul. Komandorska 118/120, 53-345 Wroclaw, Poland, pawel.rokita@ue.wroc.pl

A household model with two decision makers may be used as a good approximation of any household (comp [1] and [2]). The literature on personal finance, like life cycle financial planning, often builds on the achievements in the area of intertemporal choice theory ([3], [8], [6]), mortality models ([4], [5], [6]) and consumption models ([7], [9]) for a single person. Financial planning for households with two decision makers differs in many respects from a similar task for a single individual. For instance, the discounted expected utility function, as proposed by Yaari [9], can not be directly used. The underlying survival process is bidimensional, with four possible states, instead of two states ("is alive" or "is dead") of one-person survival process. Moreover, the consumption process built on the bivariate survival process is path dependent. It is not sufficient to know who of the household members is alive at a moment. If one person is alive and the other is not, it matters when he or she died and whether it was before or after the retirement age of that person. This is because the quantities that influence feasible consumption, like cumulated surplus and cumulated investments, depend on the history of the survival process, not only on its current state. The property of path-dependence and cumulative nature of the analyzed process makes it necessary to analyze probabilities of whole scenarios, encompassing trajectories from the start of the plan to the end of the household life, instead of conditional probabilities of survival. The second difference is the possibility of internal risk transfer within the household. This also allows to take advantage of the so called partial retirements (comp. [1]). The third is the significance of premature-death risk (in the case of a single individual, longevity risk might have financial consequences for the situation of the household, whereas early-death risk was nonexistent, but some influence on bequest potential).

To take all these characteristics of households into account, an original value function is proposed. Feldman, Pietrzyk and Rokita [2] explain in more details its assumptions, construction and technique of application in household life-long financial planning. The general formula of the function is given by eq. (1).

2 Assumptions and definitions

A two-person household is taken into consideration. It is understood as a household in which there are two decision makers, called *main household members*. They are assumed to intend to remain members of the household until their death. Other household members, like children or elderly parents of the main two are treated just as part of financial situation of the two persons. Advantages of the model with a household defined like this are discussed by Feldman, Pietrzyk and Rokita [2].

The model of household financial plan used here is based on the proposition by Feldman, Pietrzyk and Rokita [1]. It is a cumulated-surplus-based framework with a goal function of the household based on discounted expected utility of consumption and bequest. As it has already been mentioned in the introduction, for the goal function the term "value function" is used, which is in compliance with the terminology of decision making theory, when referring to extensions or generalizations of utility. It is assumed that the aim of the household financial plan is to cover all financial goals, preserve financial liquidity throughout the whole life cycle, and maximize consumption, taking into account all constraints. The value function is given with the equation (1):

$$V(c_0, v; \boldsymbol{\Psi}) = \tag{1}$$

$$= \sum_{D_2^*=E(D2)-\gamma^*}^{E(D2)+\delta^*} \sum_{D_1^*=E(D1)-\gamma^*}^{E(D1)+\delta^*} p \left[\alpha \left(\sum_{t=0}^{max\{D_1^*, D_2^*\}} \frac{1}{(1+r_c)^t} u\big(C(t; D_1^*, D_2^*)\big)\big(\gamma(t)+\delta(t)\big) \right) + \beta \frac{1}{(1+r_B)^{max\{D_1^*; D_1^*\}}} u\big(B(max\{D_1^*, D_2^*\}; D_1^*, D_2^*)\big) \right] \longrightarrow max$$

where:

$u(.)$ – utility function (the same in all segments of the formula),

c_0 – at the moment 0,

v – proportion of Person 1 investment in joint one-period contribution of the household ($v \equiv v_1$, $v_1 = 1 - v_2$),

$\boldsymbol{\Psi}$ – information about investment program mix used for financing household goals; for each program belonging to the mix the following information is provided:

- $\Psi_{i,1}$ – contribution to a program i,
- $\boldsymbol{\Psi}_{i,2}$ – vector of dates of withdrawals from program i (since the program may be used to finance more than one goal, more than on withdrawal needs to be allowed for),
- $\boldsymbol{\Psi}_{i,3}$ –vector of withdrawal sums corresponding to withdrawal dates $\boldsymbol{\Psi}_{i,2}$,

β – bequest motive parameter ($\beta = 1 - \alpha$),

$E(Di)$ – expected lifetime of person I (unconditional, i.e., under the condition of the state at the moment of program start t_0),

$C(t; D_1^*, D_2^*)$ – consumption at the moment t, for the scenario in which $D1 = D_1^*$ and $D2 = D_2^*$,

$B(t; D_1^*, D_2^*)$ – bequest (cumulated investments and surplus of both household members) at the moment *t*, for the scenario in which $D1 = D_1^*$ and $D2 = D_2^*$,
γ^* – premature-death risk aversion parameter (number of years that household takes into consideration),
δ^* – longevity risk aversion parameter (number of years that household takes into consideration),
$\gamma(t)$ – premature death risk aversion measure (depends on γ^*),
$\delta(t)$ – longevity risk aversion measure (depends on δ^*),
p – probability that at least one person is alive,
r_C – discount rate of consumption,
r_B – discount rate of bequest.

Financial goals are represented in the model as cash out-flows that are planned for some dates in the future. The magnitude and accomplishment times of household goals are deterministic. Financial goals may be any planned cash outflows that can not be financed from ongoing incomes. This means that financing of the goals needs to be planned. In practice it usually means saving and investing. One of the constraints for the financial plan optimization procedure is the requirement that all goal expenditures are covered.

There is one particular goal distinguished, which is a private retirement (usu. intended to fill up retirement gap). This goal is also represented by a single cash outflow on the date of retirement, which is interpreted as the transaction of buying a life annuity at this moment. To the most typical other goals of a household belong: buying a house or financing university education for children.

It is assumed that the household uses systematic investment programs for financing the goals. It does not mean that the investment must be a formal systematic investment product offered by a financial institution. It may be also just an informal plan of systematic saving and investing by the household. To be able to work it into the life-long financial plan, the household needs to define this systematic investment program in terms of class of assets, amount of monthly (quarterly, annual) contribution and time of start and end of the program.

Financial goals may be financed by different combinations of programs. One program may be used to finance one or many goals. There are possible different financing structures. A financing structure is understood here as a general assignment of investment programs to goals, without defining concrete programs, but just showing whether a number of different goals is financed with one program or many programs (for instance, given three goals numbered 1, 2, and 3, the goal 1 and 3 may be financed by one program, and the goal 2 with another, where programs are not specified yet).

3 Financing household goals

Let the household have n goals that may be financed with m systematic investment programs. The choice of a financing structure is made on the basis of a number of criteria, but not all financing structures are directly comparable at once. This is why selecting of a financial structure is a four-step procedure.

The choice of programs may be made only within a given financing structure, because different financing structures are incomparable in respect of the main financial criteria, that is – risk and expected return. The decision about the level of risk is made earlier, before any financing structure is defined. This is the decision regarding the classes of assets that are taken into consideration. This determines also expected returns on particular assets, assuming that the market is efficient. Profitability of investment programs and financing structures built of them is, however, not equivalent to expected return of the assets used by them. The systematic investment programs may show different cost efficiency, even if they are based on the same asset class. The first criterion of choice used in this procedure is thus cost efficiency in the meaning of the height of contribution paid by the household, other parameters being common and fixed for all program mixes belonging to a given financing structure.

There are the following four steps of the choice procedure:

Step 1 – the choice of risk level

The decision about accepted level of risk is indeed an asset allocation decision. It needs to be suited to the age of household members, their preferences, investment experience and economic education.

Step 2 – creating financing structures

A general financing structure may be defined as a $(2 \times n)$ matrix $\boldsymbol{S}$, in the first row of which there are indices denoting goals, whereas the second row is constructed using the following algorithm:

a) put any symbol, say "A", in the first field of the second row of matrix $\boldsymbol{S}$:
$S_{2,1} = A$,

general way of financing goal 1 is then: $S_{:,1} = \begin{bmatrix} 1 \\ A \end{bmatrix}$;

b) for $j := 2\ to\ n$ repeat the following:

c.1) if goal j is to be financed with the same program as some of the goals considered so far (i.e., goals: $[1, \dots, j-1]$), let it be a goal v, $(1 \le v \le j-1)$, whose general way of financing is: $S_v = \begin{bmatrix} v \\ \Upsilon \end{bmatrix}$,
then substitute
$S_{2,j} = \Upsilon$,
general way of financing of goals $[1, \dots, j]$ is then: $S_{:,[1:j]} = \begin{bmatrix} 1 \cdots v \cdots j \\ A \cdots \Upsilon \cdots \Upsilon \end{bmatrix}$.

c.2) otherwise (goal j is intended to be financed with another program),
assign to goal j a symbol that has not been used yet, say Ξ; thus, substitute:
$S_{2,j} = \Xi$,
general way of financing goals $[1, \dots, j]$, is then: $S_{:,[1:j]} = \begin{bmatrix} 1 \cdots v \cdots j \\ A \cdots \Upsilon \cdots \Xi \end{bmatrix}$;

For example, if there are 3 goals and 3 programs, 4 general financing structures are possible:
$\boldsymbol{S1} = \begin{bmatrix} 1\,2\,3 \\ AAA \end{bmatrix}$, $\boldsymbol{S2} = \begin{bmatrix} 1\,2\,3 \\ AAB \end{bmatrix}$, $\boldsymbol{S3} = \begin{bmatrix} 1\,2\,3 \\ ABB \end{bmatrix}$, $\boldsymbol{S4} = \begin{bmatrix} 1\,2\,3 \\ ABC \end{bmatrix}$.
It should be stressed that sequences ABB and ACC are identical. The same refers to, for example, sequences AAB and AAC. General financing structures inform whether goals are to be financed by the same or different programs, not determining, yet, which concrete programs might there be.

Step 3 – contribution minimization

For each general financing structure separately a concrete program mix is chosen. The program mixes contain concrete programs with the information which program is destined to finance which financial goal of the household. The criterion of choice at this stage is minimization of joint contribution paid by the household. The program mix with the lowest cost for the household from amongst of mixes belonging to a given financing structure will be further referred to as *efficient mix*.

Step 4 – Maximization of value function of the household

The value function of the household is fed with efficient mixes from step 3.

For each efficient mix the value function is maximized. The decision variables in each such optimization are:

- consumption – retirement investment division (c_0), that is – the proportion of private retirement investment to joint income of the household at the moment 0.
- retirement contribution division (v_0), that is – the relative share of Person 1 in the joint retirement investment contribution of the household.

The optimization is performed under a number of constraints, to which belong, among others, the condition of sustaining financial liquidity throughout the whole lifecycle for expected trajectory of the survival process, covering all financial goals set by the household and using the mix of investment programs that is under consideration.

The program mix, despite the fact that it is an argument (set of arguments) of the value function, is not treated as decision variable at this stage. It is just given as a set of parameters, determining some constraints of optimization procedure. Then, for each of the efficient program mixes, the result of optimization is found. Levels achieved by the value function for the optimum solutions are compared. The program mix for which the maximum of the value function is the highest is chosen as globally optimal.

4 Example of application

Expected cumulated surplus trajectory of a household with some stylized characteristics (age of its members, incomes, costs of living) under the assumption that the household has no financial goals, is presented in **Figure 1**.

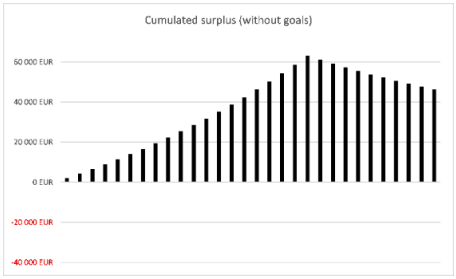


Figure 1 Expected cumulated surplus without financial goals

The incomes are sufficient to support consumption, assuming that no other expenditures are planned and that no unplanned events occur. At this stage of the research no risk is taken into consideration. The analysis is performed for expected values of any stochastic factors. When deterministic expenditures that reflect planned goals of the household are added, the expected trajectory of surplus will be certainly change. The example presented beneath illustrates the role of financing in realization of household goals.

Let us assume that the household has two financial goals. Corresponding cash out-flows are presented in **Figure 2**.

Figure 2 Financial goals

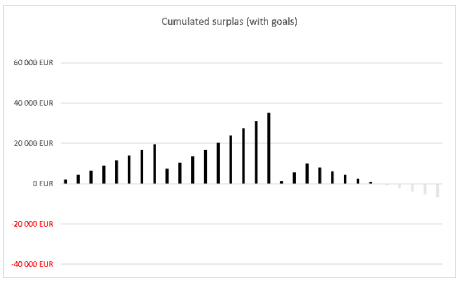


Figure 3 Expected cumulated surplus with financial goals

In the **Figure 3** there is presented expected trajectory of cumulated surplus, taking into account realization of financial goals.

As it may be red from **Figure 3**, realization of financial goals causes substantial shortfall. Moreover, this shortfall would be incurred at late phase of the life cycle. There would be, thus, little chance to find financing for it. The shortfall shows potential size of debt that would be necessary to support liquidity of the household, but in fact such a financial plan could have not been accepted at start, and the situation would not happen.

To avoid shortfall, the household members may decide to renounce some part of their consumption for the sake of investments. Let us assume that the household pays on regular basis some contribution to systematic investment programs. The question now arises, whether it should be one program financing the two goals or two separate programs, each destined to finance one goal. Assuming that for both financing structures: 1) one program to two goals and 2) two programs to two goals, an efficient program mix was selected. This means that the programs are already the least-cost ones available for a given financing structure. The results of financing the two goals with two separate and one common program are shown in **Figure 4** and **Figure 5**, respectively.

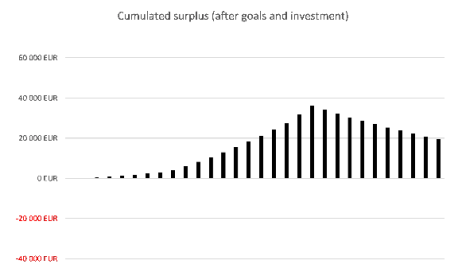


Figure 4 Financing two goals with two programs

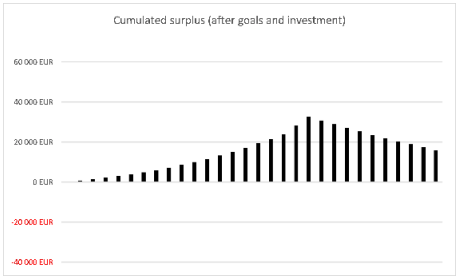


Figure 5 Financing two goals with one program

A comparison of the results of the two program mixes is provided in the **Figure 6**.

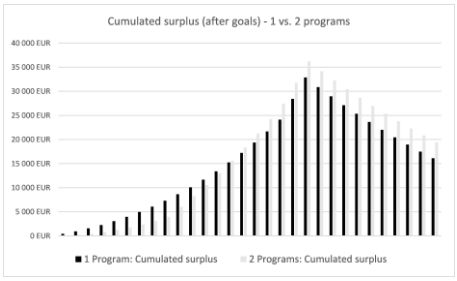


Figure 6 Comparison of two program mixes

The decision which of the two solutions (the one presented in the **Figure 4** or **Figure 5**) to choose cannot be made without additional information on household preferences. This information is coded in the value function. It contains, amongst others, the information about risk aversion, propensity to consume and bequest motive.

The program mix for which the value function takes on the higher value is chosen as the optimal. It needs to be emphasized that for each program mix an optimization procedure is performed and only the results of optimization are compared (selecting the maximum of maxima).

5 Summary

The main features of the approach used here are: the two-person household model (allowing for longevity risk and premature-death risk diversification), the use of value function in which risk aversion is expressed in terms of number of years – parameters γ^* and δ^* in eq. (1) (straightforward interpretation and application), reducing number of survival scenarios taken into consideration (comp. [2]) and a four-step procedure of financing program mix selection (sections 3 and 4). This is a result of a part of research into life-long household financial planning with multiple goals and multiple financing opportunities, taking into account preferences of the household and various types of risk it faces. Further research in this area will concentrate on augmenting the model by other types of risk than the risk related to length of life, relaxing the constraint that all financial goals must be fully accomplished (what will require considering also utility of goals) and analysis of stability of the model (particularly sensitivity of plan optimization results to small changes of some household characteristics, like income, costs of living, etc.).

Acknowledgements

The research project was financed by The National Science Centre (NCN) grant, on the basis of the decision no. DEC-2012/05/B/HS4/04081.

References

[1] Feldman, L., Pietrzyk, R., Rokita, P.: A practical method of determining longevity and premature-death risk aversion in households and some proposals of its application. In: *Data Analysis, Machine Learning and Knowledge Discovery, Studies in Classification, Data Analysis, and Knowledge Organization* (Spiliopoulou, M., Schmidt-Thieme, L., Janning, R., eds.). Springer International Publishing, 2014, 255–264.

[2] Feldman, L., Pietrzyk, R., Rokita, P.: Multiobjective optimization of financing household goals with multiple investment programs. In press. (accepted for publishing in *Statistics in Transition. New Series*), 2014.

[3] Fisher, I.: *The Theory of Interest*. Macmillan, New York, 1930.

[4] Gompertz, B.: On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*, 115 (1825), 513–585

[5] Makeham W.M. (1860): On the Law of Mortality and the Construction of Annuity Tables. Journal of the Institute of Actuaries and Assurance Magazine, 8 (1860), 301–310.

[6] Milevsky, M. A., Huaxiong, H., Salisbury, T. S.: Yaari's Lifecycle Model in the 21st Century: Consumption Under a Stochastic Force of Mortality (January 6, 2011). Available at SSRN: http://ssrn.com/abstract=1816632 or http://dx.doi.org/10.2139/ssrn.1816632.

[7] Modigliani, F., Brumberg, R. H.: Utility analysis and the consumption function: an interpretation of cross-section data. In: *Post-Keynesian Economics* (Kenneth K. Kurihara, ed.), Rutgers University Press, New Brunswick, NJ, 1954, 388–436.

[8] Samuelson, P.A.: A Note on Measurement of Utility. *The Review of Economic Studies*, 4 (1937), 155–161.

[9] Yaari, M.E.: Uncertain Lifetime, Life Insurance and Theory of the Consumer. *The Review of Economic Studies*, 32(2) (1965), 137–150.

A microeconometric analysis of household expenditures on transport services in the Czech Republic

Vítězslav Píša[1], Jan Brůha[2]

Abstract. The determinants affecting the demand for transport are of the crucial interest of policymakers, investors and other stakeholders including researchers. This paper investigates the structure of household expenditures on various types of transport. To do that, we formulate a microeconometric model based on the extended version of the Tobit model - that can be used to estimate probability that a given household spends on particular transport modes. The model is estimated in the Bayesian framework.

Keywords: Transport Demand, Censored Data, Multivariate Tobit Model

JEL classification: C24, C34, D10
AMS classification: 91G70

1 Introduction

The determinants affecting the demand for transport are of the crucial interest of policymakers, investors and other stakeholders including researchers. While policymakers generally realize large public project private investors usually need to know the rentability of their investments. The characteristics of population in particular location play therefore essential role in decision making process into which transport mode invest in the most effective way.

This paper focuses on identification of key drivers of different types of transport. We particularly ask who and under what conditions predominantly spends on given transport modes. For the sake of the analysis the extension of multivariate tobit based model is applied to cross sectional data from the Czech Household Budget Survey. Tobit types of models properly handle the issue of censored data with many zero values of dependent variables where traditional least square methods provide biased estimates. The results infer about probabilities that a given household spends on different transport mode.

The paper is organized as follows. Section 2 presents the model and its estimation. Section 2.2 describes data used for the analysis, while Section 3 presents the results. The last section concludes.

2 Model formulation

The original Tobit model is a statistical model proposed by Tobin [4] as model from limited dependent variable models (LDV) family in a latent (unobserved) variable framework.

$$y = \max(y^*, \mathbf{0}), \tag{1}$$

$$y^* = x\beta + u \tag{2}$$

where the observed variable y is equal to y^* if the latent variable is positive and zero otherwise, the latent variable y^* fulfills traditional linear model assumptions.

The Tobit model has been applied in statistics to modeling observations where a significant part takes zero values. Zero values can have various interpretations depending on the context, and one of the main

[1]Kolin Institute of Technology, Okružní 703, Kolín, Czech Republic, vitezslav.pisa@centrum.cz
[2]Kolin Institute of Technology, Okružní 703, Kolín, Czech Republic, jan_bruha@yahoo.co.uk

application is the estimation of the demand for goods with infrequent purchases or which a part of the population does not consume.

In the multivariate setting the assumed data-generating process for observation Y_{id} is:

$$Y_{id} = \max(Y^*_{id}, 0). \tag{3}$$

$$Y^*_{id} = X_{id}\beta_d + u_{id} \tag{4}$$

where X_{id} are observed characteristics for observation i and component d and β_d is the vector of parameters. The random terms u_{id} are uncorrelated across individuals, but may be correlated across components, i.e. $cov(u_{id_1}, u_{id_2}) = \sigma_{d_1,d_2}$.

The dependence between components Y_{id} can be caused by (i) the observed covariates X_{id} (i.e., similar agents can have similar characteristics that translate into similar transport behaviour) and (ii) by the correlation of unobserved terms u_{id_1} and u_{id_2} (i.e., two travel modes may be substitutes or complements). If the covariates X_{id}, parameters β_d and error terms u_{id} are suitable stacked to X_i, β and u_i, then the model (3),(4) can be compactly rewritten as:

$$Y_i = \max(Y^*_i, \mathbf{0}), \tag{5}$$

$$Y^*_i = X_i\beta + u_i, \tag{6}$$

where $\mathbf{0}$ is the D-dimensional vector of zeros, and u_i is a D dimensional error terms with a zero mean and the covariance matrix $\Sigma = (\sigma_{ij})$, and the operator max in (5) is applied component-wise. In other words, Y^* is a random vector with the multivariate normal (MVN) distribution with the mean $X_i\beta$ and the covariance matrix Σ.

Given the knowledge of β and Σ (which is the goal of the estimation), it is possible to make predictions for an individual i with observed characteristics X_i about y_i. The prediction includes the probability that the positive observation will occur (probability that $y_{id} > 0$), the expected value of y_{id} or the expected value of y_{id} conditional on $y_{id} > 0$. Also cross-statistics may be of interest, such as the conditional probability of $y_{id_1} > 0$ given $y_{id_2} > 0$ for $d_1 \neq d_2$ e.g. how the probability that an agent will use public transport changes if we know that he/she also uses individual transport.[1]

2.1 Estimation of the model

Given the latent-data structure of the model, the natural Bayesian approach is to use a Gibbs sampler. The idea is simple: if the latent variables Y^* are observed, then the Bayesian estimation of the parameters β and Σ is basically reduced to the estimation of the seemingly unrelated regression (SUR) model, and there are efficient algorithms for Bayesian estimation of the SUR model (Geweke [3]). However, if the parameters β and Σ are known, then it is possible to sample Y^* conditional on observations Y.

The Gibbs sampler thus iterates between (i) sampling of the parameters β and Σ using draws of Y^* as data; and (ii) sampling from the distribution of Y^* given actual observations Y and the current sample of the parameters β and Σ. The full sampling from the distribution of Y^* given actual observations Y and parameters would be time-consuming[2]. To reduce the computational time, one needs to realize that it is not necessary to sample from the conditional distribution of Y^*, but it is sufficient to sample from Y^* sequentially. In the paper, we provide details for the sampling from the posterior distribution for the independent normal Wishart prior, i.e., the prior on β is specified as a MVN distribution, while the prior on Σ is an inverted Wishart distribution. In what follows, I denotes the sample size (the number of units), while D denotes the dimensionality of the model (i.e., the number of components of Y_i).

[1] This cross statistics are in general hard to evaluate analytically, but may be easily simulated given the knowledge of the parameters.

[2] It is not an easy task to sample from the truncated MVN distribution. There are various possibilities of doing this: for example the obvious accept-reject algorithm, which is time-consuming if the probability of positive observation is not high. The alternative is another Gibbs sampler, which would iterate over conditional marginal densities of Y^*. Under this possibility, the main Gibbs sampler would incorporate another sampler. However, a Monte Carlo experiment (Brůha [1]) suggests that the drawing from the joint conditional distribution of Y^* do not bring any benefits in computational time or accuracy relative to the sequential sampler used in this paper.

Algorithm

- Set priors for β and Σ; the prior for β is the MVN distribution with parameters β_0 and S_0, the prior for Σ is the inverted Wishart distribution with parameters c_0 and C_0^{-1}. Set the burn-in periods N_1 and genuine iterations N_2.
- Set initial values of the parameters $\beta^{(1)}$ and $\Sigma^{(1)}$.
- Set $n = 1$ and run the following Gibbs sampler:
  1. Given $\beta^{(n)}$ and $\Sigma^{(n)}$ and Y, sample $Y^{*(n)}$ as follows:
     (a) For $i = 1, \ldots I$ do the following
     (b) For $d = 1, \ldots D$ do the following
     (c) If $Y_{id} > 0$ set $Y_{id}^{*(n)} = Y_{id}$, else derive the conditional distribution of $Y_{id}^{*(n)}$ which is the univariate normal distribution truncated at zero from the left[3] with the parameters[4] $\widetilde{\mu}_{id}$ and $\widetilde{\sigma}_{id}^2$.
  2. Given the current sample of $Y^{*(n)}$ and $\Sigma^{(n)}$, sample $\beta^{(n+1)}$ from the MVN distribution with the mean $\widehat{B}\left[\sum_{i=1}^{I} X_i'(\Sigma^{(n)})^{-1}Y_i^{*(n)} + S_0^{-1}\beta_0\right]$ and the covariance matrix $\widehat{B} = \left[\sum_{i=1}^{I} X_i'(\Sigma^{(n)})^{-1}X_i + S_0^{-1}\right]^{-1}$;
  3. Given the current sample of $Y^{*(n)}$ and $\beta^{(n+1)}$, sample $\Sigma^{(n+1)}$ from the inverted Wishart distribution with the following parameters: $c_1 = c_o + I$ and $C_1^{-1} = C_0^{-1} + \sum_{i=1}^{I}(Y_i^{*(n)} - X_i\beta^{(n+1)})(Y_i^{*(n)} - X_i\beta^{(n+1)})'$;
  4. Set $n = n + 1$ and repeat until $n = N_1 + N_2$.
- Disregard the first N_1 draws and use the remaining N_2 samples as the sample from the posterior distribution for β and Σ.

See Brůha and Brůhová-Foltýnová [2] for a more detailed description of the algorithm and for the discussion of the frequentist estimation of the model using the methods-of-moments.

2.2 Data

The data that we use come from the Household Budget Survey completed by the Czech Statistical Office and the estimation is done for the year 2010. The database contains information about 3251 respondents - households, their demographic characteristics, and structure of expenditures. The expenditures on the various types of transport are of our main interest. The main characteristics of the expenditure distribution are briefly summarized in the table 1.

In the table 1, we provide the minimum *conditional on non-zero expenditure* (Min (cond.))[5], the mean, the maximum, the probability of reporting positive expenditures (Prob), the standard deviation (STD) and the standard deviation *conditional on non-zero expenditure* (STD (cond.))[6] Obviously, the transport behaviour differs significantly among households, and the standard deviation are large.

Figure 1 then displays histograms for expenditures of households on transport modes, for which they reported positive expenditures.

[3]The sample from the **univariate** truncated distribution is simple, e.g., the inverse method gives $\widetilde{\mu}_{id} + \widetilde{\sigma}_{id}\Phi^{-1}(u\Phi(-\widetilde{\mu}_{id}/\widetilde{\sigma}_{id}))$, where Φ is the cumulative distributive function of the standard normal distribution, Φ^{-1} is the quantile function of the standard normal distribution, and u is a draw from the uniform random variable on the interval [0, 1].

[4]The parameters of the conditional distribution are given as follows: $\widetilde{\mu}_{id} = \mu_{id}^{(n)} + \Sigma_{d,-d}^{(n)}(\Sigma_{-d,-d}^{(n)})^{-1}(Y_i^{*(c)} - \mu_{i,-d}^{(n)})$, $\widetilde{\sigma}_{id}^2 = \Sigma_{d,-d}^{(n)} - \Sigma_{d,-d}^{(n)}(\Sigma_{-d,-d}^{(n)})^{-1}\Sigma_{-d,d}^{(n)}$, where $\mu_{id}^{(n)}$ is the dth element of the unconditional mean $X_i\beta^{(n)}$, $\Sigma_{d,-d}^{(n)}$ is the dth row of the unconditional covariance matrix $\Sigma^{(n)}$, $\Sigma_{-d,-d}^{(n)}$ is the submatrix of $\Sigma^{(n)}$ after deleting the dth row and dth column, and $\mu_{i,-d}^{(n)}$ is created from the unconditional mean $X_i\beta^{(n)}$ by deleting the dth row. Finally, $Y_i^{*(c)}$ contains the current sample from Y_i^* and its elements are given as $Y_{i\delta}^{*(c)} = Y_{i\delta}^{*(n+1)}$ for $\delta < d$ and $Y_{i\delta}^{*(c)} = Y_{i\delta}^{*(n)}$ for $\delta > d$.

[5]In other words the minimum in the sample of strictly positive values. It does not make sense to report the unconditional minimum – as there are non-users for each transport group, the minimum would be trivially zero.

[6]In other words the standard deviation in the sample of strictly positive values.

Transport mode	Min (cond.)	Mean	Max	Prob	STD	STD (cond.)
Motor fuels	25	4785	96226	0.74	5677	5721
Combined transport	2.3	574	1523	0.41	1335	1783
Buses (public city)	2.0	570	6591	0.33	195	862
Buses (inter city)	3.2	585	22537	0.59	1386	1683
Taxi	2.2	25	4997	0.08	190	594.1
Rail	3.6	329	28904	0.45	1078	1506
Airplane	4.6	76	29996	0.01	959	5947
River and sea	2.5	7	1900	0.02	75	3387

Table 1 Transport Expenditures (in CZK per household member, 2010 prices)

Explanatory Variable	Posterior mean	2.5 % HPD	97.5 % HPD
Income	1186.79	1089.01	1301.62
Persons	1021.37	811.40	1207.76
EA Persons	3283.82	2597.16	3998.51
Unemployment	-803.56	-1536.33	-11.28
Big cities	-7033.08	-7877.90	-6193.02
Cities	-5466.39	-6307.96	-4547.77
Villages	-3933.76	-4838.14	-2994.35

Table 2 Motor Fuels

Explanatory Variable	Posterior mean	2.5 % HPD	97.5 % HPD
Income	-8.46	-34.40	17,60
Persons	-13.37	-59.93	31.22
EA Persons	115.71	-45.91	274.63
Unemployment	121,82	-59.63	280.20
Big cities	-381.11	-572.04	-195.54
Cities	-895.72	-1087.00	-695.07
Villages	-980.43	-1177.56	-770.61

Table 3 Buses: Public City Transport

Explanatory Variable	Posterior mean	2.5 % HPD	97.5 % HPD
Income	-6.59	-45.43	33.49
Persons	-26.74	-86.13	37.37
EA Persons	789.55	540.31	1027.46
Unemployment	46.95	-190.28	270.52
Big cities	-925.03	-1192.56	-639.19
Cities	138.98	-157.65	415.02
Villages	179.66	-91.22	506.61

Table 4 Buses: Inter City Transport

3 Estimation results

We applied the above described model to the Czech household data. In particular, we investigate the determinants of expenditures on the four selected transport modes (motor fuels, two types of bus transport, and rails) given 7 socio-demographic characteristics. Tables 2 – 5 contain the estimation results. For each parameter we provide the posterior mean, and 2.5% and 97.5% quantile of the high probability density (HPD) region of the posterior distribution, which is a rough analog of the frequentist's confidence

Explanatory Variable	Posterior mean	2.5 % HPD	97.5 % HPD
Income	68.41	30.74	105.36
Persons	58.07	-11.56	127.25
EA Persons	130.96	-124.17	378.21
Unemployment	-115.41	-347.73	114.27
Big cities	-1111.80	-1375.64	-867.52
Cities	-1107.40	-1386.40	-824.38
Villages	-1376.94	-1695.81	-1105.70

Table 5 Rails

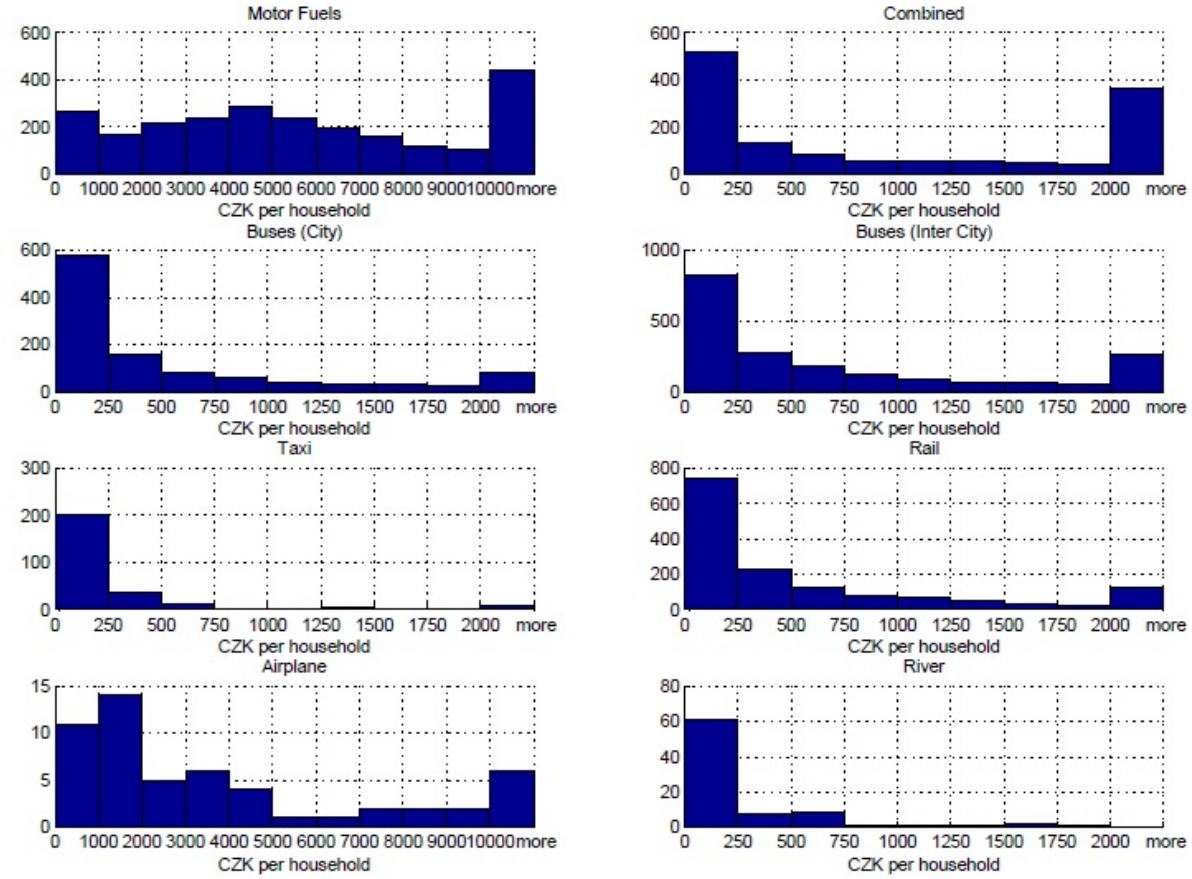


Figure 1 Histograms of transport expenditures

intervals.

We considered the following socio-economic variables: the household income (normalized to the living minimum standard), the number of persons, the share of economically active persons in the household, the duration of unemployment spell of the household head (as fraction of the year), and three dummies related to where the household lives: big cities (over 50 000), cities (5000-50000), and villages (less than 5000). Note that we do not include the constant into the model, and therefore we are allowed to use the full set of these dummies.[7]

Our results suggest that income strongly increases expenditures on motor fuels and also on rails, but rather decreases expenditures for two types of bus transport. Therefore with growing income one can substitute bus transport with, for example, private car transport. The number of persons and the share of economically active persons increases only expenditures on motor fuels. This can be partly explained, for example, by the fact that sharing private car becomes economically feasible with more persons. The

[7]This means that the effect of these dummies should be considered relative to each other.

unemployment spell of the household head (even after controlling for income) decreases expenditures on all transport modes with possible exception of buses used for public city transport that could also stem from the fact that unemployed person substitutes private car transport with public transport. Living in big cities decreases expenditures on motor fuels, but increases expenditures on rails (again after controlling for income), and the opposite holds for living in villages.

4 Conclusion

In this paper, we analyzed the transport behaviour of Czech household. We ask the question of determinants of expenditures on transport modes, we presented the statistical model based on the multivariate Tobit model and its statistical estimation using the Gibbs sampler. We find that certain-socio economic characteristics of households are strong predictors of expenditures on transport behaviour. The results can be also useful for formulating more elaborated transport demand models in our future research.

Acknowledgements

The research on this paper has been supported by the grant of the Czech Scientific Foundation *No. 14-22932S*. The support is gratefully acknowledged.

References

[1] Brůha J.: *A Comparison of Bayesian Methods for Estimating of Multivariate Models with Censored Data*. KIT mimeo, 2014.

[2] Brůha J., and Brůhová-Foltýnová, H.: *Active Travel in Prague: Assessing Heterogeneity in the Population*. Working paper, Kolin Intitute of Technology, available at http://www.koltech.cz/english/data/Active-travel-Prague.pdf, 2013.

[3] Geweke, J.: *Contemporary Bayesian Econometrics and Statistics.* Wiley Series in Probability and Statistics. John Wiley and Sons, New York, 2005.

[4] Tobin, J.: Estimation of relationships for limited dependent variables. *Econometrica* **26(1)** (1958), 24–36.

DSGE model with government sector and inflation targeting: Structural changes of Czech economy in a period of recession

Vratislav Pisca [1], Osvald Vašíček[2]

Abstract. In the paper we compare the structure of the Czech economy in two periods: before and during the Great Recession. This small open economy is represented by two variations of a medium scale nonlinear dynamic stochastic model of a general equilibrium with fiscal and monetary policy. This model framework is based on a Portuguese approach introduced by Almeida (2009). We decided to use this model, because it captures the important features of the Czech economy in a reasonable detail. Since this concept was primarily constructed for the estimation of an economy in a currency union, we had to add a Taylor type monetary policy and a nominal exchange rate into the model. Both model variations are estimated with the use of bayesian techniques with the same priors, in order to analyse structural changes linked to the Great Recession. With the use of the shock decomposition we will analyse the impact of the exogenous shocks on the development of the Czech output. Especially, we will focus on the effects of the fiscal policy in the period of the Great Recession and the European debt crisis.

Keywords: nonlinear DSGE model, Great Recession, European debt crisis, Czech economy, structural changes, fiscal policy, inflation targeting.

JEL classification: 91B64
AMS classification: E32, E58

1 Introduction

The Czech economy experienced a very calm period with an uninterrupted growth of the gross domestic product since the 1997 currency crisis. But in the year 2007 the financial crisis broke out and later transformed into the Great Recession that striked the Czech economy as well. Between 2008Q3 and 2009Q3 the real GDP decreased by aproximately 5%. A recovery of the Czech economy is very slow and it seems that the GDP finally reached its potential in 2013Q4. This paper is aimed to identify and analyze the structural changes of the Czech economy linked to the development of the Great Recession.

For this purpose we use a small open economy DSGE model, whose structure is based on a Portuguese approach introduced in Almeida (2009). We chose this concept, because both economies are small and very open, and so they are very strongly affected by foreign fluctuations. This medium-scale model concept can describe all main features of an economy.

Since Portugal is a member of the Eurozone and it does not have its own currency, we had to modify several model equations and add the nominal and real exchange rates in line with Gali (2008) and Justiniano, Preston (2010). Moreover we had to add the Taylor type monetary policy with inflation targeting proposed by Senaj (2010), because the Czech National Bank performs its own independent monetary policy unlike the Portuguese central bank takes the nominal interest rate from the ECB. This monetary policy is forward looking and adjusts the nominal interest rate according to the difference of the expected inflation from the inflation target and to the deviations of the gross domestic product from its potential.

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, 385260@mail.muni.cz
[2]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, osvald@econ.muni.cz

2 Model

Since, we mainly want to examine the development and structural changes of the Czech economy, we decided to rather elaborate the model framework developed by Almeida (2009), that was designed for the Portuguese economy. Both these economies are small and open. Structure of the model is quite standard, therefore, we will describe only the most important features of the economy represented by model. The model contains six general parts: households, firms, aggregators, government, central bank and foreign sector.

2.1 Households

The households maximize their discounted lifetime utility, which is positive in difference of consumption and consumption habit and negative in labour supply. The households maximize their lifetime utility by choosing a level of real consumption, real investment, (next period) domestic bonds holdings and (next period) foreign bonds holdings subject to a budget constraint and a capital accumulation equation. An utility is affected by a consumption shock ε^c and an inverse labour supply shock ε^l. We assume that the households own the capital stock. Capital follows an accumulation equation and it is determined by the investment made by the households. Investment is subject to adjustment costs S, where these costs are positive function of the changes of the investments between current and previous period. A value of capital in the next period is affected by an investment shock ε^i. The households receive resources from: the domestic and foreign bonds, which yield the domestic nominal interest rate and risk-adjusted foreign nominal interest rate respectively. The foreign bonds risk-premium is a decreasing function of the real stationary holdings of foreign assets of the entire domestic economy and an increasing function of the risk-premium shock ε^ϕ. The households also obtain resources from: participating in the market of state-contingent securities, from government in the form of transfers and collect profits of firms in the form of dividends. The households supply differentiated labour. This fact gives the households some market power and enables them to charge additional markup for the supplied labour. This markup is affected by a wage markup shock μ^w. A Calvo-type wage rigidity is assumed in the model (Calvo parameter ξ^w). The households that cannot reoptimize their wage only update their wage in accordance with the current period inflation target (affected by a shock in inflation target $\varepsilon^{\bar{\pi}}$),previous period inflation rate and current period growth rate of the permanent technology shock ζ.

2.2 Firms

In the model, there are intermediate and final good firms. Intermediate good firms are of three types: domestic, import and composite good firms. Four types of final good firms are assumed: consumption, investment, government consumption and export firms.

- Intermediate goods firms: Domestic good firms rent capital, and hire labour to produce heterogeneous domestic goods. The domestic intermediate goods are produced by the Cobb-Douglas production function, which is positively affected by a temporary technology shock ε^a . The domestic good firms hire labour and rent capital in perfectly competitive markets and take the wage and rental rate of capital as given.
- Import goods firms: Import goods firms buy a homogenous foreign goods and differentiate them into import goods by a brand naming.
- Composite good firm: There is one composite good firm that combines the homogeneous domestic good and import good to produce homogeneous composite good via CES production function. Both input and output markets where the composite good firm operates are perfectly competitive, and thus, the firm takes all prices as given.
- Final goods firms: There are four types of final goods firms: private consumption, investment, government consumption and export. Final goods firms buy a certain amount of composite goods and differentiate them by a brand naming to produce heterogeneous final goods. All input markets of final good firms are perfectly competitive and the firms take the input price as given. Calvo-type rigidities of final good prices are assumed in the model. If a firm cannot reoptimize its price, it

only adjusts the price to the current inflation rate target and to the previous period inflation of its respective good.

- All kinds of firms except composite good firm operate on monopolistically competitive markets of their respective output and this fact allows them to charge an additional markup μ^f, $f \in \{d, m, c, i, g, x\}$ over the marginal costs and to generate profits. This markups are affected by their respective shocks ε^{μ^f}.

2.3 Aggregators

The mismatch between the supply of differentiated products and labour and the demand for homogeneous products and labour is solved by aggregators. Labour and each type of differentiated product have one respective aggregator that buys all the different varieties and combines them to produce homogeneous product via the CES technology. Each aggregator operates in a perfectly competitive market of its input and output, therefore, it takes prices as given.

2.4 Goverment

The government obtains its resources from consumption and income taxes, that are affected by their respective shocks ε^c and ε^l, and via issuance of domestic bonds, and spends the resources on government consumption goods, transfers to the households and payment of debt services. The government consumption is affected by a shock ε^g. In order to prevent explosive debt path, a fiscal rule is assumed. Debt to GDP ratio cannot permanently exceed the target value $b = (B/GDP)^{tar}$. If this situation occurs, transfers automatically decrease in order to decrease the government deficit and satisfy the fiscal rule in the long-run. The taxes and government consumption are given exogenously as shocks.

2.5 Central bank

We extended the original model of Almeida (2009) with the domestic monetary authority, which allows us to model the independent monetary policy of the Czech National Bank. In line with Senaj (2010), monetary authority sets the nominal interest rate in accordance with Taylor type interest rate rule in order to reach the inflation target and keep the output gap closed. Furthermore, the Taylor rule contains a smoothing parameter ρ^{mp}, which leads to a gradual adjustment of the interest rate. Deviations from the Taylor rule are modelled as discretionary monetary policy shocks ε^{mp}. The monetary rule has a following form:

$$R_t = \rho_{mp} R_{t-1} + (1 - \rho_{mp})[\phi^{\pi}_{mp}(\pi_{t+1} - \bar{\pi}_t) + \phi^{gdp}_{mp} \hat{gdp}_t] + \varepsilon^{mp}_t \tag{1}$$

2.6 Foreign sector

The foreign variables (output, nominal interest rate and inflation) are assumed to be exogenously given as AR(1) processes with their respective shocks ε^{y^*}, ε^{r^*} and ε^{π^*}. Moreover, we assume that there is a foreign permanent technology shock ζ^* . The foreign demand for domestic export goods depends on the nominal exchange rate, the share of foreign and domestic price levels, foreign output and foreign elasticity of substitution between domestic export good and foreign good. Nominal exchange rate is assumed to adjust the income from foreign bonds so as to fulfil the uncovered interest parity i.e. the income from domestic bonds must be equal to the income from foreign bonds expressed in domestic currency.

3 Parameterisation of model

3.1 Data

The time series of seventeen observed variables are used for the estimation of the model. The Czech economy is described by: gross domestic product (gdp), consumption (C), investment (I), government

consumption (G), export (E), import (M), nominal interest rate (r), real wage (w), employment (L), inflation target ($\bar{\pi}$) and three types of inflation: GDP inflation (π^d), consumption good inflation (π) and investment good inflation (π^i). The foreign sector is represented by the time series of gross domestic product (y^*), CPI inflation (π^*) and nominal interest rate (r^*) of 17 Eurozone countries. CZK/EUR real exchange rate (RER) is used to link the domestic economy with the foreign sector. Fifteen time series were obtained from the Eurostat. Real exchange rate was obtained from the ECB database and the inflation target was extracted from the CNB annual reports. All time series contain quarterly data from the period 1999Q1 to 2013Q4, of total length of 60 observations. Each time series was stationarised with use of the H-P filter with $\lambda = 1600$. We divided the dataset into the before-crisis period of 1999Q1-2008Q3 containing 39 observations and the during-crisis period of 2008Q4-2013Q4 containing 21 observations.

3.2 Calibration

The same calibration and priors were used for both estimations. We decided to calibrate 11 parameters. The depreciation of capital δ was set to 0.02, which corresponds to an annual depreciation rate of 8%. The growth of the permanent technology ζ was set to 1.005, which implies an annual potential output growth of 2%. The steady state inflation π was set to 1.005, which is consistent with inflation goal 2% in the European Union. The discount rate β was set to 0.999 to produce a steady-state long-run nominal interest rate of 4.5%. Target debt to GDP ratio was set in line with the Czech fiscal policy to $b_{tar} = 0.6$. Two calibrations of parameters were obtained from data. Steady-state government to output ratio g_y was set to 0.43, and the share of domestic exports in EA17 output ω_{for} was set to 0.005. The rest of the parameters were set in line with Almeida (2009): Consumption tax ($\tau_c = 0.304$), income tax ($\tau_l = 0.287$), productivity of capital ($\alpha_d = 0.323$) and the fiscal rule parameter ($d_g = 0.9$).

Parameters		
	Longname	Value
δ	Depreciation	0.02
β	Discount rate	0.999
α_d	Capital productivity	0.323
ω_{for}	Share of export on foreign output	0.005
π	Inflation	1.005
g_y	Realtive government expenditure	0.430
τ_c	Consumption tax	0.304
τ_l	Income tax	0.287
ζ	Technology	1.005
d_g	Fiscal rule	0.90
b_{tar}	Fiscal rule	0.60

Table 1 Calibration

3.3 Estimation

As was mentioned above, we performed two separate estimations, with the same calibration and priors. The only difference was in the dataset used for each estimation. Estimation was carried out in matlab with use of a dynare toolbox, version 4.3.3. Two chains of Metropolis-Hastings algorithm with 1.000.000 replications each and drop rate set to 0.6 were generated. Table 2 presents the posterior estimates of structural parameters of both estimations.

4 Results

The most considerable change occurs in the parameter of domestic intermediate goods markup μ^d. This parameter decreased in the crisis, which means that a willingness to substitute is higher and terms on market are tighter for producers of heterogeneous domestic goods in the Czech economy. This fact is in line with a situation during the crisis. Next interesting change occurs for a parameter of an adjustment cost of investments S, which shifts from the value 7.58 to the value 8.83. This means, that an implementation of the investments is more difficult and a cost of it is higher. The parameter of a habit in a consumption of households h declined from 0.72 to 0.67 which reflects worse sentiments of the households during the

recession. Since we analyze the structural changes, we skip the posterior results of shocks, but we want to point out, that a significant change occurred to the shock of the foreign output. The value of its AR parameter decreased from 0.95 to 0.87 and the standard deviation of its inovation increased from 0.24 to 0.69.

Parameters		*Priors*			*Posterior Before*		*Posterior During*	
	Long name	Type	Mean	stdev	Mean	stdev	Mean	stdev
ϑ_h	Elasticity, import	Norm	1.000001	0.50	0.4447	0.0835	0.5089	0.0897
ϑ_{for}	For. elasticity export	Norm	1.50	0.50	1.3396	0.3672	1.0546	0.4069
ω_h	Import share	Beta	0.32	0.05	0.1827	0.0218	0.1968	0.0244
σ_l	Labour elasticity	InvG	2.00	0.50	1.9206	0.5056	1.9077	0.4389
h	Consumption habit	Beta	0.64	0.05	0.7223	0.0421	0.6714	0.0464
μ^w	Wage markup	InvG	1.25	0.10	1.3261	0.0683	1.3113	0.0668
μ^d	Domestic markup	InvG	1.50	0.35	1.9040	0.2446	1.5046	0.2182
μ^m	Import markup	InvG	1.20	0.10	1.1205	0.0839	1.1087	0.0793
μ^c	Consumption markup	InvG	1.20	0.025	1.1880	0.0245	1.1884	0.0246
μ^i	Investment markup	InvG	1.05	0.05	1.0690	0.0499	1.0549	0.0501
μ^g	Government markup	InvG	1.05	0.05	1.0313	0.0476	1.0434	0.0494
μ^x	Export markup	InvG	1.5	0.10	1.5186	0.0984	1.4516	0.0913
S	Adjustment cost	Norm	7.70	1.50	7.5263	1.2169	8.5198	1.3060
ξ_w	Calvo, wage	Beta	0.83	0.10	0.7542	0.1003	0.6128	0.1378
ξ_d	Calvo, domestic	Beta	0.75	0.10	0.7029	0.0486	0.6826	0.0586
ξ_m	Calvo, import	Beta	0.50	0.05	0.3509	0.0727	0.4091	0.0750
ξ_c	Calvo, consumption	Beta	0.60	0.10	0.8028	0.0323	0.8180	0.0387
ξ_i	Calvo, investment	Beta	0.60	0.10	0.7933	0.0305	0.7318	0.0453
ξ_g	Calvo, government	Beta	0.60	0.10	0.5916	0.0562	0.6919	0.0626
ξ_x	Calvo, export	Beta	0.60	0.10	0.7149	0.1307	0.6739	0.1122
χ	Elasticity of risk premium	Beta	0.035	0.015	0.4050	0.0817	0.4683	0.0943
κ_w	Indexation, wage	Beta	0.50	0.10	0.5450	0.0905	0.4506	0.0969
κ_d	Indexation, domestic	Beta	0.50	0.10	0.4463	0.0989	0.4899	0.1011
κ_m	Indexation, import	Beta	0.50	0.10	0.5352	0.0957	0.5599	0.0975
κ_c	Indexation, consumption	Beta	0.50	0.10	0.6144	0.0845	0.4878	0.0968
κ_i	Indexation, investment	Beta	0.50	0.10	0.4347	0.0962	0.5228	0.1021
κ_g	Indexation, government	Beta	0.50	0.10	0.4200	0.0998	0.4608	0.1002
κ_x	Indexation, export	Beta	0.50	0.10	0.0462	0.0193	0.0260	0.0130
ϕ_{mp}^{π}	Taylor rule inflation	Gamma	1.80	0.15	1.6885	0.1393	1.8350	0.1496
ϕ_{mp}^{gdp}	Taylor rule gdp	Gamma	0.20	0.05	0.2025	0.0476	0.1623	0.0441
ρ^{mp}	Taylor rule smoothing	Beta	0.80	0.10	0.8513	0.0276	0.8660	0.0295

Table 2 Priors and posteriors of the structural parameters

4.1 Shock decomposition

In the Figure 1 a historical shock decomposition of the Czech gross domestic product is shown. The shocks are divided into six groups for a better clarity. Group "Rest of the shocks" contains the permanent and temporary technology shocks and a shock in risk premium, group "Prices" contains markups shocks of all aggregators, group "Preferences" contains consumption shock, investment shock and labour supply shock, group "Monetary policy" contains only a shock in Taylor rule, group "Fiscal policy" contains shock in a government consumption and in the consumption and income taxes and group "Foreign" contains foreign permanent technology shock and the shocks in the main foreign macroeconomic variables (output, nominal interest rate and inflation). In the period before crisis, it was preferences, that had the most positive effect. Also, the foreign and monetary policy shocks had positive effects in the product. The fiscal policy was restrictive. Negative effects of a whole foreign sector occurred in the initial three quarters of the Great Recession. We can see a positive effect of the fiscal policy in period of 2008Q4 -2011Q1. At this time the debt crisis in the Europe began and government started to consolidate its budget. This affected the Czech gross domestic product negatively in the period of 2011Q2 - 2012Q4.

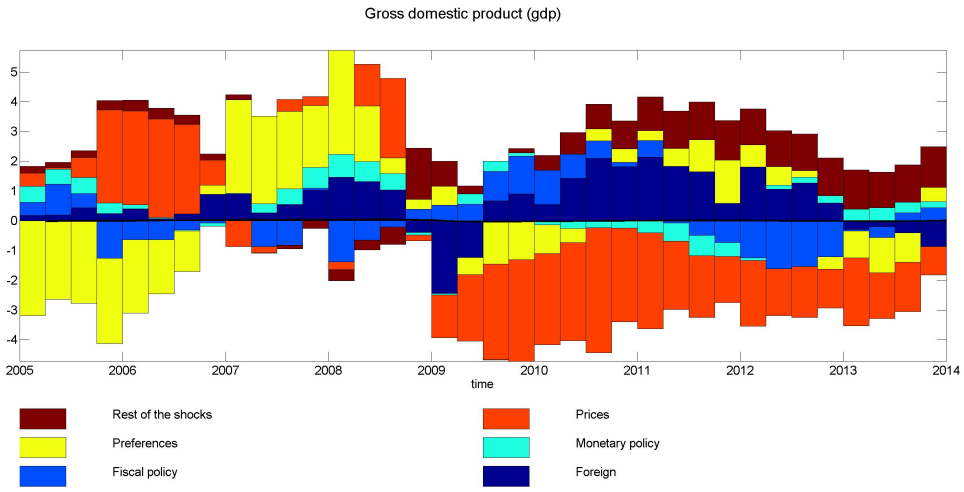


Figure 1 Shock decomposition of GDP

5 Conclusion

In the paper we analysed the structural changes of the Czech economy between the periods before and during the Great Recession. For this purpose, we estimated two medium scale models of a small open economy, which were estimated on a dataset of the Czech economy in the period before the Great Recession (1999Q1 - 2008Q3) and in the period during the Great Recession (2008Q4 - 2013Q4).

We analysed the posterior estimates of the structural parameters to discover the most significant changes of the structure of the Czech economy. We found out that the most significant changes occurred in the elasticity of substitution of the domestic goods, in the adjustment cost of investments and in the consumption habit of households. Next, we used the historical shock decomposition of the gross domestic product to describe the shocks, that affected the performance of the Czech economy, before and during the Great Recession. We found out, that before crisis it were the preference, foreign and price shocks, which had the positive effects and the fiscal policy had the main negative effect. Foreign shocks had negative effects on the domesic product in the initial three quarters of the crisis and positive effects later on during crisis. The fiscal policy was expansive during crisis until 2011Q2, when the debt crisis began and government started to consolidate its budget.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0808/2013.

References

[1] Almeida, V.: Bayesian Estimation of a DSGE Model for the Portuguese Economy. *Bank of Portugal Working Papers Series* **14/2009**, (2009)

[2] Gali, J., *Monetary Policy, Inflation, and the Business Cycle: an Introduction to the New Keynesian Framework,* Princeton University Press, (2008)

[3] Justuniano, A. and Preston, B.: Monetary Policy and Uncertainty in an Empirical Small Open Economy Model *mimeo Department of Economics, Columbia University*, (2006).

[4] Senaj, M., Výškrabka, M., and Zeman, J.: MUSE: Monetary Union and Slovak Economy model. *NBS Working paper series*, **1/2010**, (2010).

On testing the similarity of multivariate populations' structures

Dominika Polko[1]

Abstract. Comparison of populations is one of the most crucial problems in statistical inference. The most common comparisons apply to two populations, but comparisons of k populations, where $k > 2$ are also carried out. Parametric methods allow to compare the means, variances or proportions, non–parametric methods, in turn, allow to compare the distributions of two or more populations. Parametric tests require the assumptions about the distribution of the characteristic in the population to be fulfilled. Non–parametric tests do not require this assumption, but the power of these tests is usually lower.
This paper applies alternative approach to identification of differences between populations that uses permutation tests. These tests do not need fulfill the assumption about conformity with normal distribution, however have power similar to parametric tests.
The problem of comparison multivariate structures based on data in contingency tables was analyzed in the paper. The permutation tests were applied in the multivariate data structure comparison. Analysis included Monte Carlo study.

Keywords: comparing structures, multivariate data, permutation tests.

JEL Classification: C12, C15, C3, C46
AMS Classification: 62H15, 62H17

1 Introduction

One of the crucial problems in statistical inference is comparison of populations. The most common comparisons apply to two populations, but comparisons of k populations, where $k > 2$ are also carried out. The problem of comparison multivariate structures based on data in contingency tables was analyzed. The permutation tests were applied in the multivariate data structure comparison. Proposed modified classical approach for assessing the similarity of populations' structure. This paper limited discussion to three–dimensional contingency tables. Analysis included Monte Carlo study.

2 Comparing populations

There can be distinguished a wide variety of methods to compare the population. Such comparisons include both parametric and non–parametric tests. To determine if any significant differences exist in the two populations based on random sampling methods statistical inference is used. Parametric tests require the assumptions about the distribution of the characteristic in the population to be fulfilled. The most commonly statistical tests used in practice allow a comparison of the populations' parameters may include the t test or U test for comparison of the means in the two populations, one-way analysis of variance test for comparing the means in three or more populations, equality tests of two or more variance, equality test of proportions or equality test of the correlation coefficients.

In situation when the assumption about conformity with normal distribution is not fulfilled, appropriate non-parametric tests should be used. The most important tests to compare the population may include U Mann–Whitney test, Wald–Wolfowitz test and Kruskal–Wallis test. Besides comparisons of parameters, it is often necessary to refer to the comparison of distributions in the two populations. In this case, Kolmogorov–Smirnov test is used most often. In this test statistic refers to the comparison of the empirical distribution functions.

3 Permutation tests

Widely used parametric tests require the assumptions about the distribution of the characteristic in the population to be fulfilled. Non–parametric tests do not require this assumption, but the power of these tests is usually lower. Alternative approach is to use permutation tests in statistical research. These tests do not need fulfill the assump-

[1] University of Economics in Katowice, Department of Statistics, ul. Bogucicka 14 40-226 Katowice, Poland, dpolko@gmail.com

tion about conformity with normal distribution, however have power similar to parametric tests. Test procedure that uses permutation test where exact statistic's distribution is approximated by distribution obtained with the use of computer simulation can be performed [5].

The idea of these methods was proposed by R. A. Fisher in the thirties of the XXth century [3]. The essence of permutation tests is to determine test statistic and then to evaluate the empirical distribution of this statistics for each permutation of variable. When calculations affect large number of permutations, Monte Carlo study is applied. Permutation tests use an algorithm similar to that used in the bootstrap method, with the difference that in the permutation tests to determine the distribution of statistic, random sampling without replacement is used and bootstrap method use random sampling with replacement [2]. After value of statistic $\hat{\theta}_0$ is determined, N permutations are performed and values $\hat{\theta}_i$ (i = 1, 2, …, N) are determined.

The decision concerning a verified hypothesis is made on the basis of *ASL* (*achieving significance level*) value

$$ASL = P\left(\hat{\theta} \geq \hat{\theta}_0\right), \tag{1}$$

for which estimation is obtained on the basis of

$$ASL \approx \frac{card\{\, i : \hat{\theta}_i \geq \hat{\theta}_0 \,)}{N}. \tag{2}$$

This notation applies, where the H_0 rejection area is right–sided. In the case of left–sided rejection area in above notation inequality sign should be changed. When *ASL* is lower than the assumed level of significance α, then H_0 is rejected in favor of hypothesis H_1.

Research and applications related to permutation tests have increased in the recent years. Several books have recently been dedicated to these methods, for example [1][4]. In one of the research directions the development of this technique for multidimensional problems is considered.

4 Comparing populations' structures

To compare the structures of two or more populations or one population for different time periods is to calculate the appropriate measure of compliance characterized by the degree of similarity (ω) of such structures. The closer are the components of two populations' structures, the more similar are these populations' structures. As the degree of dissimilarity of the two structures S_1 and S_2 increases, the index $\omega(S_1,S_2)$ aim at zero. When compared structures are the same, the value of index $\omega(S_1,S_2)$ is 1. In literature, the following characteristics of similarity measure structures are given (see [6])

$$\omega(S_1,S_2) = 1 \Leftrightarrow S_1 = S_2, \tag{3}$$

$$\omega(S_1,S_2) = \omega(S_2,S_1), \tag{4}$$

$$\omega(S_1,S_2) \in \langle 0,1 \rangle. \tag{5}$$

To compare the structure, the structure similarity index is frequently used. The classic approach to similarity structures can be represented as follows

$$\omega = \sum_{i=1}^{k} min(\, w_i^{(s)} \,), \tag{6}$$

where $w_i^{(s)} = \frac{n_i}{\sum_{i=1}^{k} n_i}$ and s (s = 1, 2) is a population' structure number.

The similarity index that is constructed by basing it on indicators of the structure has values of the interval $\langle 0,1 \rangle$. Overview of the measure of similarity or dissimilarity of populations' structure can be found in [7][8].

5 Comparing multivariate populations' structures

Sometimes there is a need to compare the multivariate structures in two or more populations. In this case a modification of similarity index can be used for structure comparison. The next section presents a proposal of comparing multidimensional structures based on data presented in two contingency tables.

5.1 Multidimensional contingency tables

The data composed of two nominal variables usually is presented in a contingency table. Contingency tables are arrays of non–negative integers that arise from the crossclassification of a sample or a population of N objects based on a set of categorical variables of interest. The entries n_{ij} (i = 1,2,…,r, j = 1,2,…,c) are the counts for every two–way combination of rows and columns. Table 1 presents the model for such kind of data where $n_{i\bullet} = \sum_{j=1}^{c} n_{ij}$, $n_{\bullet j} = \sum_{i=1}^{r} n_{ij}$ and $n = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} = \sum_{i=1}^{r} n_{i\bullet} = \sum_{j=1}^{c} n_{\bullet j}$.

Row variable	Column variable				Row sums
	y_1	y_2	…	y_c	
x_1	n_{11}	n_{12}	⋯	n_{1c}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	⋯	n_{2c}	$n_{2\bullet}$
…	…	…	…	…	…
x_r	n_{r1}	n_{r2}	⋯	n_{rc}	$n_{r\bullet}$
Column sums	$n_{\bullet 1}$	$n_{\bullet 2}$	⋯	$n_{\bullet c}$	n

Table 1 Contingency table

A multidimensional contingency table is one that contains data of three or more variables. The multidimensional contingency table evaluated in this section represents an example of multidimensional table with three variables, with each variable being comprised of three levels. In a three–dimensional table, one of the variables is designated as the row variable, a second variable is designated as the column variable, and the third variable is designated as the layer variable. Zar [10] employs the term *tier variable* to designate the third variable. In a three–dimensional table there is a total of $r \times c \times l$ cells, where r represents the number of row categories, c the number of column categories, and l the number of layer categories (see [9]).

As noted earlier, it is possible to have a three–dimensional table that summarizes the data for three variables. Table 2 presents a multidimensional table of this type.

Row variable	Column variable													$\sum_x$
	y_1				y_2				…	y_c				
	z_1	z_2	…	z_l	z_1	z_2	…	z_l	…	z_1	z_2	…	z_l	
x_1	n_{111}	n_{112}	…	n_{11l}	n_{121}	n_{122}	…	n_{12l}	…	n_{1c1}	n_{1c2}	…	n_{1cl}	$n_{1\bullet\bullet}$
x_2	n_{211}	n_{212}	…	n_{21l}	n_{221}	n_{222}	…	n_{22l}	…	n_{2c1}	n_{2c2}	…	n_{2cl}	$n_{2\bullet\bullet}$
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
x_r	n_{r11}	n_{r12}	…	n_{r1l}	n_{r21}	n_{r22}	…	n_{r2l}	…	n_{rc1}	n_{rc2}	…	n_{rcl}	$n_{r\bullet\bullet}$
$\sum_z$	$n_{\bullet 11}$	$n_{\bullet 12}$	…	$n_{\bullet 1l}$	$n_{\bullet 21}$	$n_{\bullet 22}$	…	$n_{\bullet 2l}$	…	$n_{\bullet c1}$	$n_{\bullet c2}$	…	$n_{\bullet cl}$	n
$\sum_y$	$n_{\bullet 1\bullet}$				$n_{\bullet 2\bullet}$				…	$n_{\bullet c\bullet}$				

Table 2 Three-dimensional contingency table

Thus, $n_{i..}$ represents the number of observations in the i^{th} row, $n_{.j.}$ represents the number of observations in the j^{th} column and $n_{..k}$ represents the number of observations in the k^{th} layer. Concurrently $n_{ij.}$ represents the number of observations in the i^{th} row and j^{th} column, $n_{i.k}$ represents the number of observations in the i^{th} row and k^{th} layer and $n_{.jk}$ represents the number of observations in the j^{th} column, and k^{th} layer. The entries n_{iik} (i = 1,2,…,r, j = 1,2,…,c, k=1,2,…,l) are the counts for every three-way combination of rows, columns and layers.

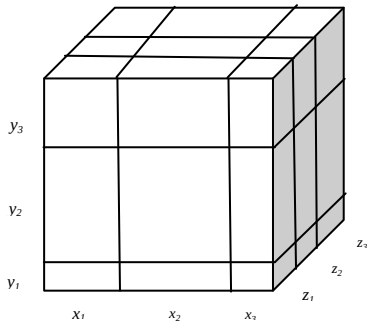


Figure 1 Three–dimensional contingency table

This paper limits discussion to three–dimensional contingency tables with each variable being comprised of three levels. Table can be represented by cube $\langle 0,1\rangle^3$, where structure of respective variants of variable (marginal structure) for each coordinate (variables *X*, *Y*, *Z*) is determined on $\langle 0,1\rangle$ (Figure 1). Moreover, the more variables that are included in the study, it becomes more difficult to test.

5.2 Comparing two multivariate populations' structures

In the next part of this article the problem of comparing two populations will be considered. Let us assume that data from the samples is given in two three–dimensional contingency tables. Let us compare the structure of the populations based on the data samples given in these contingency tables. In the case of data presentation in a three–dimensional contingency table, to compare the structures based on the data in two contingency tables the index (7) can be used

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{l} \min_{s}(w_{i\bullet\bullet}^{(s)}) \cdot \min_{s}(w_{ij\bullet}^{(s)}) \cdot \min_{s}(w_{ijk}^{(s)}), \tag{7}$$

where $w_{i\bullet\bullet}^{(s)} = \frac{n_{i\bullet\bullet}^{(s)}}{n^{(s)}}$, $w_{ij\bullet}^{(s)} = \frac{n_{ij\bullet}^{(s)}}{n_{i\bullet\bullet}^{(s)}}$, $w_{ijk}^{(s)} = \frac{n_{ijk}^{(s)}}{n_{ij\bullet}^{(s)}}$ and *s* (s = 1, 2) is a contingency table number.

The equation (7) indicates that to compute minimum observed frequency of i^{th} row of two considered contingency tables, then multiply it by minimum observed frequency of i^{th} row and j^{th} column of two considered contingency tables and then multiply it by minimum observed frequency of i^{th} row, j^{th} column and k^{th} layer of two considered contingency tables. This amounts to calculate the surface area of $(r \cdot c \cdot l)$ cuboids, where each dimension (row, column, layer) is designated observed frequency.

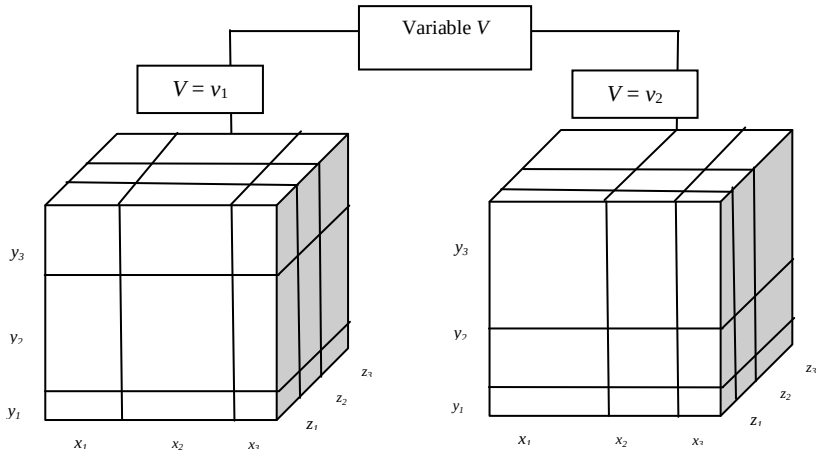


Figure 2 Representation of the four–way contingency table

To compare populations' structures based on the data in two contingency tables (classification variables *X*, *Y* and *Z*) let us consider a fourth classifying, dichotomous variable *V*, which takes the value of v_1 for the elements of the first table and the value of v_2 for the elements of the second table. The data could be written in the form

presented in Figure 2. In general, it is possible to compare *s* contingency tables. Then the variable that identifies contingency tables will take the values $v_1, v_2, \ldots, v_s$.

Data stored in two contingency tables can be written in four columns (see table 3), where the fourth column identifies the number of the contingency table.

X	***Y***	***Z***	***V***
x_1	y_1	z_1	v_1
x_1	y_1	z_1	v_1
…	…	…	…
x_r	y_c	z_l	v_1
x_1	y_1	z_1	v_2
…	…	…	…
…	…	…	v_2
x_r	y_c	z_l	v_2

Table 3 The form of the data from two three-dimensional contingency tables

The problem of comparing two contingency tables comes down to investigate the effect of variable *V* (contingency table index) on variables *X, Y* and *Z*. The hypothesis statements to test can be written $H_0 : T = 1$. Hypothesis H_0 says that the structures of the two contingency tables are identical and is verified toward hypothesis H_1 which is the negation of H_0. Distribution of statistic (7) is unknown, hence to make the decision, the permutation test will be used.

The permutation test procedure that is used for the verification of the hypothesis on the compliance of contingency tables structures is as follows

1. Assume the level of significance α.
2. Calculate the value of statistic for the sample data.
3. Perform the permutation of variable *V N*–times, then calculate the statistics test value.
4. On the basis of empirical distribution of statistic, the *ASL* value is determined. If $ASL < \alpha$, then H_0 is rejected, otherwise H_0 hypothesis cannot be rejected.

6 Monte Carlo study

The properties of the *T*–statistic described above have been analyzed in the Monte Carlo study. Three versions of compared structures were analyzed:

a) the same structures;
b) similar structures;
c) different structures.

Theoretical structures of compared populations are presented in Figure 3.

a)	$V = v_1$									$V = v_2$								
	$Y = y_1$			$Y = y_2$			$Y = y_3$			$Y = y_1$			$Y = y_2$			$Y = y_3$		
	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$
$X = x_1$	0,02	0,04	0,02	0,03	0,06	0,02	0,01	0,03	0,02	0,02	0,04	0,02	0,03	0,06	0,02	0,01	0,03	0,02
$X = x_2$	0,03	0,06	0,03	0,06	0,13	0,08	0,03	0,05	0,03	0,03	0,06	0,03	0,06	0,13	0,08	0,03	0,05	0,03
$X = x_3$	0,01	0,03	0,01	0,04	0,06	0,02	0,02	0,04	0,02	0,01	0,03	0,01	0,04	0,06	0,02	0,02	0,04	0,02

b)	$V = v_1$									$V = v_2$								
	$Y = y_1$			$Y = y_2$			$Y = y_3$			$Y = y_1$			$Y = y_2$			$Y = y_3$		
	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$
$X = x_1$	0,02	0,04	0,02	0,03	0,06	0,02	0,01	0,03	0,02	0,01	0,03	0,04	0,03	0,06	0,02	0,01	0,03	0,02
$X = x_2$	0,03	0,06	0,03	0,06	0,13	0,08	0,03	0,05	0,03	0,03	0,06	0,03	0,06	0,13	0,08	0,03	0,05	0,03
$X = x_3$	0,01	0,03	0,01	0,04	0,06	0,02	0,02	0,04	0,02	0,01	0,03	0,01	0,04	0,06	0,02	0,02	0,04	0,02

c)	$V = v_1$									$V = v_2$								
	$Y = y_1$			$Y = y_2$			$Y = y_3$			$Y = y_1$			$Y = y_2$			$Y = y_3$		
	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$	$Z = z_1$	$Z = z_2$	$Z = z_3$
$X = x_1$	0,02	0,04	0,02	0,03	0,06	0,02	0,01	0,03	0,02	0,02	0,04	0,05	0,03	0,05	0,02	0,04	0,03	0,02
$X = x_2$	0,03	0,06	0,03	0,06	0,13	0,08	0,03	0,05	0,03	0,03	0,06	0,03	0,04	0,11	0,08	0,03	0,05	0,03
$X = x_3$	0,01	0,03	0,01	0,04	0,06	0,02	0,02	0,04	0,02	0,01	0,05	0,01	0,02	0,03	0,02	0,04	0,04	0,02

Figure 3 Structures of compared populations

In simulation analyses, samples of a distribution were generated (Figure 3) of count n = 200, 500, 1000, 1500. On the basis of obtained data, hypothesis H_0 on identity of compared distributions was verified using the above described permutation test. In the analyses the significance level α = 0,05 was assumed. Estimated probabilities of rejection H_0 are presented in table 4.

T – statistic				
n	200	500	1000	1500
a)	0,054	0,037	0,060	0,045
b)	0,218	0,345	0,230	0,325
c)	0,740	0,925	0,975	1

Table 4 Estimated probabilities of rejection H_0

In case a) probability of rejection the hypothesis on identity of distributions is near α. In case c) with increasing sample size hypothesis H_0 is rejected with probability near 1.

7 Concluding remarks

The paper deals with the problem of comparing population structures. It presents alternative approach to identification of similarity of populations' structures that uses permutation tests. The procedure that enables comparing multidimensional structures on the basis of the data stored in contingency tables is used. The *T*–statistic is proposal of modification the classical approach to assessing the similarity of populations' structure. Due to the lack of information on the theoretical distribution of considered statistic, in order to compare the results the permutation test was used. These tests do not need fulfill the assumption about conformity with normal distribution. This paper limited discussion of three–dimensional contingency tables.

References

[1] Basso D., Pesarin F., Salmaso L., Solari A.: *Permutation Tests for Stochastic Ordering and ANOVA*, Springer Science + Business Media, Heidelberg 2009.

[2] Domański Cz., Pruska K.: *Nieklasyczne metody statystyczne*, PWE Warszawa 2000.

[3] Efron B., Tibshirani R.: *An Introduction to the Bootstrap*, Chapman & Hall, New York, 1993.

[4] Good P. I.: *Permutation Tests: A practical guide for testing Hypotheses*, Springer-Verlag, New York, 1994

[5] Kończak G.: *O pewnym teście dla weryfikacji hipotezy o równości wartości przeciętnych w k populacjach*, In: Rola informatyki w naukach ekonomicznych i społecznych, Zeszyty Naukowe 8, tom 2. WSH Kielce, 2008, 337 – 344.

[6] Kukuła K.: *Przegląd wybranych miar zgodności struktur*, Przegląd Statystyczny, vol. 4 XXXIII (1986), 385-401.

[7] Kukuła K.: *Statystyczne metody analizy struktur ekonomicznych*, Wydawnictwo Edukacyjne, Kraków 1996.

[8] Panek T.: *Statystyczne metody wielowymiarowej analizy porównawczej*, Wydawnictwo Szkoły Głównej Handlowej w Warszawie, Warszawa 2009.

[9] Sheskin D.J.: *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall, Boca Raton, 2003.

[10] Zar, J. H.: *Biostatistical analysis* (4th ed.), Upper Saddle River, NJ: Prentice Hall, 1999.

The Comovement of Financial and Trade Integration: Wavelet Co-Spectrum Approach

Jitka Poměnková[12], Zuzana Kučerová[3]

Abstract. The paper is focused on the identification of comovement between financial and trade integration in the EU member countries over the period 1993-2012 using a wavelet approach. We use yearly data of selected indicators of financial and trade integration. Our findings show a significant comovement for all countries and indicators in long cycles in period 2000-2009. We also found the comovement between frequencies corresponding to the business cycle in 1993-1994, 2003-2004 and in general 2008-2010. The process of financial integration was stronger in the "old" EU member countries (plus Cyprus and Malta) and the process of trade integration in the "new" member countries. We conclude that there was a strong relation between financial and trade integration in the pre-crisis period (before 2008) in the EU countries compared to period after 2008.

Keywords: wavelet co-spectrum, comovement, financial integration, foreign trade integration

JEL Classification: C18, C8, E44, F36, F42
AMS Classification: 62P20, 62M10, 91B84

1 Introduction

The integration of financial markets is an integral part of the overall economic integration; it can also support an economic growth because it helps to remove the exchange rate risk and the barriers in the cross-border capital movement. Financial integration also increases the financial markets efficiency (Baele *et al*. [4]). There was another argument for greater financial market integration: the common monetary policy (in monetary unions), which should be implemented through the common financial system. It should be efficient and functioning well to provide a predictable, smooth, and effective transmission of the monetary policy. Mundell [26] also supported the capital mobility as one of the main criteria for sustainable fixed exchange rate arrangements. Financial integration and international diversification of assets help to reduce the risk of economic recession in the case of negative macroeconomic shocks. Nevertheless, the process of financial integration of the past decade before 2007/2008 financial crisis was associated with the accumulation of risks and the national and supranational financial regulation and supervisory practices lagged behind the highly integrated, fast expanding and sophisticated financial sector (European Commission [10]). As a consequence, it is important to monitor and measure the state of integration in all segments of financial markets.

Trade integration is one of the most important factors which have an impact on the financial integration. Lane and Milesi-Ferreti [22], [23] specify three important linkages between foreign trade and trade with foreign assets and liabilities. Firstly, trade with goods and services evokes the corresponding financial transactions; international trade and international financial flows are thus able to equilibrate the balance of payment. Secondly, a high share of bilateral trade linkages between countries leads to a higher portion of financial transactions; investors have a better knowledge of foreign companies from these countries and are thus more prone to buying the shares of these companies (the "familiarity effect"). And thirdly, a high degree of trade openness reflects the liberal approach of macroeconomic policy authorities in these countries both in the area of foreign trade and cross-border capital flows.

The aim of the paper is to assess the comovement between financial and trade integration in the selected EU country groups in 1993-2012. We use quantity-based measures of financial integration[4] derived from the countries' international investment positions with a view to the foreign trade. The paper is structured as follows. In

[1] Faculty of Economics, VSB-Technical University of Ostrava, Dept. of Economic Policy, Sokolská tř. 33, Ostrava 1, 701 21, jitka.pomenkova@vsb.cz.
[2] Faculty of Electrical Engineering and Communication, Brno University of Technology, Dept.of Radio Electronics, Technická 12, 616 00 Brno, pomenkaj@feec.vutbr.cz
[3] Faculty of Economics, VSB-Technical University of Ostrava, Dept. of Economic Policy, Sokolská tř. 33, Ostrava 1, 701 21, zuzana.kucerova@vsb.cz.
[4] For the classification of financial integration indicators se e.g. Baele *et al*. [4].

section two we provide short literature review from methodological as well as economic point of view. In section three, the method, data, countries and indicators of financial and trade indicators are defined. The fourth section contains the results of the wavelet analysis. The last section summarises the results and brings conclusions.

2 Literature Review

Many interesting issues are posed by synchronization of economic time series. Analysis in the time domain predominated for long time. It was used several instruments such as correlation analysis and its modification. Extension of the standard correlation analysis proposes the cointegration, common features (Engle and Kozicki [9] or common cycles and codependence (Vahid and Engle [30]). Unfortunately only time domain perspective does not provide a sufficient answer to the question how to quantify the degree of synchronisation and how to analyse the evolution of such a synchronisation during time with respect to the occurrences in the economic time series. A great attention was paid to these techniques in connection to the European integration process and the theory of optimum currency area. A number of studies were written to assess the readiness and alignment of countries to enter the Eurozone (Darvas and Szapáry[8]). Consequent methodological approaches proceeded to techniques in frequency domain based on spectral and cross spectral analysis in the past several decades. Therefore, the comovement can be viewed from the perspectives of a dynamic correlation and phase shift methods, coherence or a squared coherency. Croux et al. [7] provide the theoretical background with a practical application on business cycles in Europe and the USA. Fidrmuc et al. [11] use dynamic correlation to estimate the determinants of output comovements among OECD countries. As wrote Iacobucci and Noullez [16], this method allows a detailed study of the comovement of time series. Current methodological approaches, also used in this study, interconnect both time and frequency domain in one. There are several time-frequency approaches such as a wavelet analysis, a time-varying AR process or a time-varying periodogram (Blumenstain *et al.* [5]). The combination of both domains provides a more efficient means of statistical analysis. Rua [27] measures the co-movement among Germany, France, Italy, and Spain via the cross-wavelet spectrum. The same approach is extended to the wavelet power spectrum, the wavelet cross-spectrum and to wavelet coherence and significance tests by Jiang and Mahadevan [17]. Aguiar-Conraria and Soares [2] use a wavelet analysis to study business cycle synchronisation across the EU-15 and the Euro-12 countries. Fidrmuc *et al*. [12] use wavelet spectrum analysis to study globalization and business cycles in China and G7 countries. The advantage of the wavelet analysis is that it can capture the features of non-stationarity time series due to the simultaneous time-frequency decomposition of inputs (Jiang and Mahadevan [17]).

One of the area for application such methods is the analysis of changes in financial integration with respect to trade integration. Lane and Milesi-Ferretti [21] discuss a suitable data set and the development of methodology how to estimate foreign assets and liabilities for a large set of both industrial and developing countries. Lane and Milesi-Ferretti [23] study international balance sheets to examine the relation between foreign assets and liabilities on one side and a set of various repressors (trade openness, GDP per capita, external liberalisation, financial depth, stock market capitalisation, privatisation revenues etc.) on the other side. They find that international trade and stock market capitalisation are the two most important variables influencing international balance sheets. This study was updated in Lane and Milesi-Ferretti [25]. As stated in Spiegel [29] trade integration intensifies financial integration and thus a common currency fosters the foreign trade of the euro area countries (the "euro effect"). Sebnem *et al*. [28] find that the impact of euro effect on financial integration is primarily driven by eliminating the currency risk. While financial and trade integration are highly correlated processes, trade in goods does not play a key role in explaining the positive effect of euro on financial integration. Kose *et al*. [18] focus on the roles of trade and financial integration in driving the growth-volatility relationship. They find that both trade and financial integration significantly weaken the negative association between output volatility and growth. Kose *et al*. [19] analyse the impact of selected macroeconomic variables (the depth of financial markets, trade openness, real GDP per capita, macroeconomic policies stability, institutional quality, and the regulation of an economy) on the level a country financial openness (the sum of financial assets and liabilities relative to nominal GDP). They deduce that foreign direct investments and cross-border flows of equity securities are safer for the economy than cross-border flows of debt securities especially in the case of a low level of a country's financial openness. Complementarity between bilateral trade in goods and bilateral asset holdings in a simultaneous gravity equations framework is studied by Aviat and Coeurdacier [3]. According to their findings, trade activities in goods and assets are closely related however the impact of asset holdings on trade in goods is smaller. The same results are confirmed in Kucerova [20] using the simultaneous equation model. Aizenman and Noy [1] construct a theoretical framework leading to two-way feedbacks between financial and trade openness and then verify these feedbacks empirically. According to them countries cannot choose the degree of financial openness independently of their degree of trade openness.

Presented studies do not consider country groups in the sense of the “new” and “old” EU countries (plus Cyprus and Malta). Therefore, our paper focuses on the relation between financial and trade integration process in

the EU member countries and also separately in the EU16 and EU10 countries. In this manner we can distinguish between the processes in these groups which are quite different and are thus worth examining.

3 Methods

The continuous wavelet transform (CWT) of time series *s(t)* is defined as

$$S_{CWT}(a,b) = \int_{-\infty}^{\infty} s(t)\frac{1}{\sqrt{b}}\psi\left(\frac{t-a}{b}\right)dt, \quad b>0, a\in R, \tag{1}$$

where *a* is the time position (time shift), *b* is the parameter of dilatation (scale) of the mother wavelet $\psi(\cdot)$. The dilatation is related to the Fourier frequency and numerator of the fraction $\sqrt{b}$ ensures the conservation of energy. The CWT transforms input time series from the time representation to the time-scale domain. As Jiang and Mahadevan [17] denoted, the wavelet transform provides more natural results, because it uses shorter windows for higher frequencies compare to the Fourier transform. To be the inversible transform, mother wavelets must be mutually orthogonal, have zero mean value and limited to finite time interval. That is

$$i)\ \int_{-\infty}^{\infty}\psi(t)dt=0,\quad ii)\ \int_{-\infty}^{\infty}\psi^2(t)dt=1,\quad iii)\ 0<C_\psi=\int_0^{\infty}\frac{|\hat{\psi}(\omega)|^2}{\omega}d\omega<\infty;\quad \hat{\psi}(\omega)=\int_{-\infty}^{\infty}\psi(t)e^{-i\omega t}dt, \tag{2}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(\omega)$. To satisfy assumptions for the time-scale analysis, waves must be compact in time and frequency representation as well.

The analysis of the relation between two time series can be measured via several quantities (Croux *et al.* [7]). In the frequency domain we can use dynamic correlation, the coherence or cohesion. Similar analysis could be also done in the time-scale domain. In our approach we use the wavelet cross spectrum between two inputs $s_1(t)$ and $s_2(t)$ for their time-scale representation $S_{CWT,1}(a,b)$ and $S_{CWT,2}(a,b)$ calculated according to the formula (1) which can be defined as (Jiang and Mahadevan [17])

$$S_{12} = S_{CWT,1}(a,b)S_{CWT,2}(a,b). \tag{3}$$

4 Data

We use yearly data 1993-2012 to calculate the measures of financial integration. The data are drawn from the International Monetary Fund International Financial Statistics online database (IMF [13]). Incomplete data for some countries and some years have been completed from the on-line database External Wealth of Nations Mark II (Lane and Milesi-Ferretti [24]). Data concerning nominal exports and imports (in USD) are extracted from the on-line database IMF IFS. Data concerning nominal GDP (in USD) are extracted from the on-line database IMF World Economic Outlook (WEO) Database (IMF [14]).

Each indicator used in the wavelet analysis (see Table 1) was calculated as the average of corresponding values for 26 representative EU countries (EU26). In order to separate and compare the processes of financial and trade integration in the developed countries and formerly central-planned economies, we divide the EU countries into two subsamples: EU16 (Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Malta, Netherlands, Portugal, Spain, Sweden, and the United Kingdom) and EU10 (Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, and Slovenia). Indicators are derived from the international investment position and its structure and defined in Lane and Milesi-Ferretti [23]. Figure 1 illustrates time series of the financial and trade integration indicators used in our analysis.

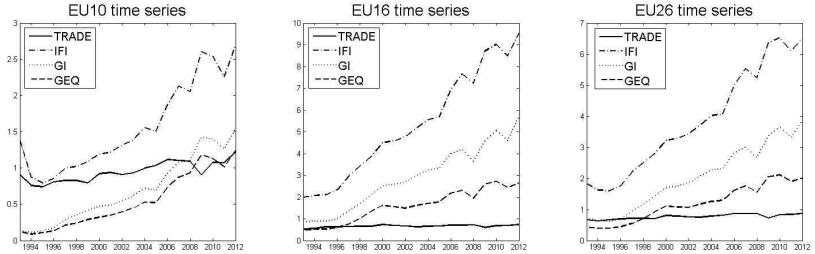


Figure 1 Time series representation of financial indicators

Source: IMF [14], [15], Lane and Milesi-Ferretti [24]

Abbreviation	Short notation	How to calculate
TRADE	Indicator of trade integration	$TRADE_{it} = \frac{(EX_{it} + IM_{it})}{GDP_{it}}$, EX_{it} is the total sum of exports of country i in time t, IM_{it} is the total sum of imports of country i in time t and GDP_{it} is the nominal GDP of country i in time t.
IFI	Quantity-based measure indicator of financial integration	$IFI_{it} = \frac{(FA_{it} + FL_{it})}{GDP_{it}}$, FA_{it} is the stock of total foreign assets of country i in time t, FL_{it} is the stock of total financial liabilities of country i in time, GDP_{it} is the nominal GDP of country i in time t.
GI	Investment-based measure of financial integration	$GI_{it} = \frac{(FDIA_{it} + FDIL_{it} + PEQA_{it} + PEQL_{it} + PDEA_{it} + PDEL_{it})}{GDP_{it}}$, $FDIA_{it}$ is the stock of foreign direct investment assets of country i abroad, $FDIL_{it}$ is the stock of foreign direct investment liabilities of the rest of the world in country i, $PEQA_{it}$ is the stock of portfolio equity assets of country i abroad, $PEQL_{it}$ the stock of portfolio equity liabilities in country i, $PDEA_{it}$ the stock of portfolio debt assets of country i abroad, and $PDEL_{it}$ is the stock of portfolio debt liabilities in a country i.
GEQ	Equity-based measure of financial integration	$GEQ_{it} = \frac{(FDIA_{it} + FDIL_{it} + PEQA_{it} + PEQL_{it})}{GDP_{it}}$

Table 1 Indicators description

5 Results

For the comovement analysis in the time-scale domain we use wavelet analysis especially wavelet co-spectrum. From the family of mother wavelet functions we used the Morlet wavelet (Gencay *et al*. [13]). The results are presented in charts for EU10 figures 2-4a, for EU16 figure 2-4b and for EU26 figure 2-4c. With respect to the chart representation and for better visibility we present on the x-axis time (yearly range 1993-2013), on the y-axis specific periods (2-64 periods: from short cycles to long cycles) and on the z-axis the value of co-spectrum.

In all country groups and for all indicators, the wavelet co-spectrum shows several areas of comovement bounded by the time and frequency. The most significant common features arise for long cycles (64-32 quarters) for all analysed indicators (*TRADE* versus *IFI*, *GI* and *GEQ*). For EU16 and EU26, this comovement occurs in the same period (2000-2009). The EU10 case shows similar results; for the *IFI* indicator it is also in period 2000-2009, in the case of *GI* and *GEQ* the time period with significant comovement is only slightly shorter (2001-2009).

Originally, the business cycle frequency was proposed by Burns and Mitchel [6] with frequency band stated between 6 and 32 quarters. The second area with common features occurs in the part of business cycle frequency range, especially between 18 and 32 quarters for all countries in period 1993-1994 (*IFI* and *GI*) and 1993-1995 (*GEQ*) more or less weak. In some cases – EU16 (*IFI*), EU26 (*IFI*, *GI*, *GEQ*) for 2002-2003 – we can identify an additional comovement in business cycle frequencies.

The last area with comovement is in the middle cycles which can be also taken as medium business cycles in frequency range 12-20 quarters). In EU26 it is between 2008-2010 for all indicators (*IFI*, *GI*, *GEQ*). In EU16, it is in 2009-2010 for *GI* and *GEQ*, for *IFI* it happened in 2008-2011. The EU10 case has different results among indicators; 2008-2010 (*IFI*), 2008-11 (*GI*) and 2007-2011 (*GEQ*).

In our analysis, the long cycles can be viewed as a time series trend because data are not detrended. As the wavelet analysis allows an application on non-stationary data (Jiang and Mahadevan [17]) and with respect to the short sample size, we do not provide detrending of inputs. Detrending or usage of appropriate filter techniques is possible but with respect to the small sample size we have to admit the noising of results. We can assume the wavelet analysis applied on detrended data can provide better information about short cycles (under the level of 12 quarters), but in case of 18 quarters we worry about quality of filtered data. Therefore, we skipped this methodological step in our contribution.

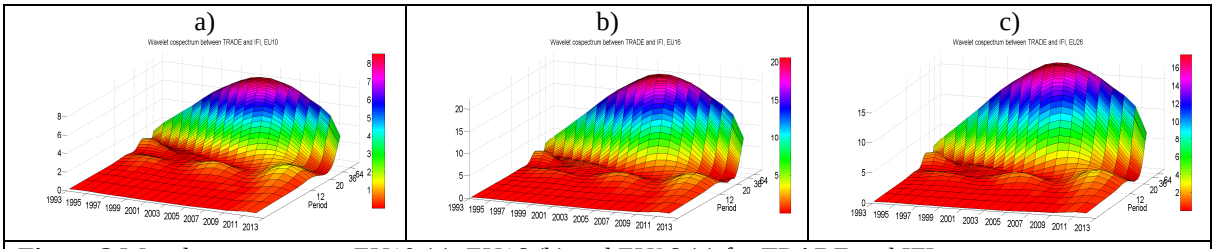


Figure 2 Wavelet co-spectrum EU10 (a), EU16 (b) and EU26 (c) for TRADE and IFI

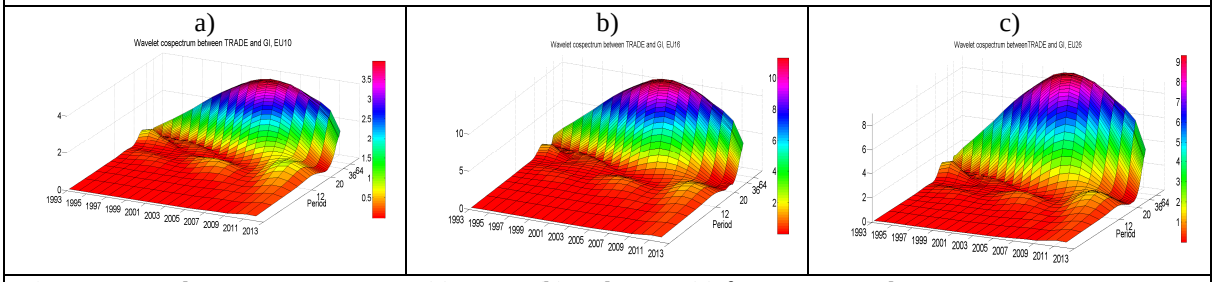


Figure 3 Wavelet co-spectrum EU10 (a), EU16 (b) and EU26 (c) for TRADE and GI

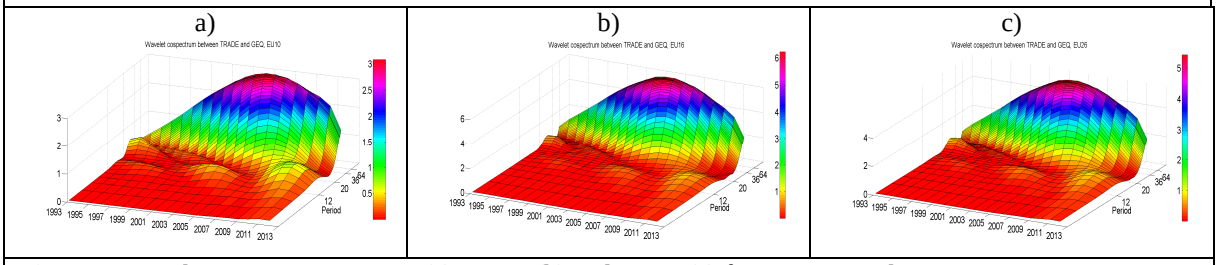


Figure 4 Wavelet co-spectrum EU10 (a), EU16 (b) and EU26 (c) for TRADE and GEQ

Source: Authors' calculations

6 Conclusion

Financial integration is a substantial condition for introducing a single currency and a common monetary policy in such an area. Nevertheless, financial integration can be strongly influenced by the trade integration, i.e. foreign trade is an important factor influencing financial integration. The aim of the article was to assess the relation between two types of economic integration (financial and trade integration) in the EU member countries over the period 1993-2012. For achieving this aim, the method of wavelet analysis is used.

On the basis of the analysis in the time-frequency domain, we can confirm the existence of dependency for long cycles (over 32 quarters) in 2000-2009. Additionally, we can see existence of comovement in 1993-1994 for business cycle frequencies. Subsequently, we found an additional cycle comovement in 2002-2003 especially in the EU16 countries and generally after crisis in 2008 and ended mostly in 2010. To sum up, the two integration processes are considerably interconnected (mainly in the long term), i.e. the more the countries trade the more financially integrated they are. It was proved in our empirical analysis, i.e. it confirms a strong relationship between the two main balance-of-payment components: the current account and the financial account.

Acknowledgements

This work was supported by the VSB-Technical University of Ostrava, Faculty of Economics under Grant SGS SP2014/115.

References

[1] Aizenman, J., and Noy, I.: Endogenous Financial and Trade Openness. *Review of Development Economics* **2** (2009), 175–189.

[2] Aguiar-Conraria, L., Soares, M. J.: Business cycle synchronization and the Euro: A wavelet analysis. *Journal of Macroeconomics* **33** (2011), 477–489.

[3] Aviat, A., and Coeurdacier, N.: The Geography of Trade in Goods and Asset Holdings. *Journal of International Economics* **1** (2007), 22–51.

[4] Baele, L., Ferrando, A., Hördahl, P., Krylova, E., and Monnet, C.: Measuring European Financial Integration. *Oxford Review of Economic Policy* **4** (2004), 509-530.
[5] Blumenstein, J., Poměnková, J., Maršálek, R.: Comparative Study Of Time- Frequency Analysis Approaches With Application To Economic Indicators. In: *Proceedings of the 26th European Conference on Modelling and Simulation ECMS 2012* , Univ Koblenz Landau, Koblenz, Germany, 2012, 291-297.
[6] Burns, A.F., Mitchell, W.C. (1946): Measuing Business Cycles. New York, National Bureau of Economic Research, pp.590, ISBN: 0-870-14085-3;
[7] Croux, C., Forni, M., and Reichlin, L.: A Measure of Comovement for Economic Variables: Theory and Empirics. *The Review of Economics and Statistics* **2** (2001), 232–241.
[8] Darvas, Z., Szapary, G.: Business Cycle Synchronization in the Enlarged EU. *Open Econ Review* **19** (2008), 1–19.
[9] Engle, R. F, Kozicki, S.: Testing for common features. *Journal of Business & Economic Statistics* **11** (1993), 369-380
[10] European Commission: European Financial Stability and Integration Report 2010. *Commission Staff Working Paper* **SEC(2011) 489** (2011).
[11] Fidrmuc, J., Ikeda, T., and Iwatsubo, K.: International transmission of business cycles: Evidence from dynamic correlations. *Economics Letters* **114** (2012), 252-255.
[12] Fidrmuc, J., Korhonen, I., Poměnková, J.: China and the World economy: Wavelet spectrum analysis of business cycles. Forthcomming in *Applied Economic Letters* (2014).
[13] Gençay, R., Selçuk, F., Whitcher, B.: *An introduction to wavelets and other filtering methods in finance and economics.* Academic Press, California, 2002.
[14] International Monetary Fund: *International Monetary Fund International Financial Statistics* [online]. 2013a [cit. 2013-12-27]. Available from Internet: http://www.imf.org/external/data.htm.
[15] International Monetary Fund: *World Economic Outlook* [online]. 2013b [cit. 2013-12-27]. Available from Internet: http://www.imf.org/external/data.htm.
[16] Iacobucci, A.: Spectral Analysis for Economic Time Series. *Documents de Travail de l'OFCE* **2003-07** (2003).
[17] Iacobucci, A., Noullez, A.: A Frequency Selective Filter for Short-Length Time Series. *Computational Economics* **25** (2005), 75–102.
[18] Jiang, W., Mahadevan, S.: Wavelet spectrum analysis approach to model validation of dynamic systems. *Mechanical Systems and Signal Processing* **25** (2011), 575–590.
[19] Kose, M. A., Prasad, E. S., and Terroned, M. E.: How Do Trade and Financial Integration Affect the Relationship between Growth and Volatility? *Journal of International Economics* **1** (2006), 176–202.
[20] Kose, M. A., Prasad, E. S., and Taylor, A. D.: Thresholds in the Process of International Financial Integration. *Journal of International Money and Finance* **1** (2011), 147-179.
[21] Kucerova, Z. Measuring Financial Integration in Central Europe through International Investment Positions. *Eastern European Economics* **4** (2009), 25-41.
[22] Lane, P. R., and Milesi-Ferretti, G. M.: The External Wealth of Nations: Measures of Foreign Assets and Liabilities for Industrial and Developing Countries. *Centre for Economic Policy Research Discussion Paper* **2231**(1999).
[23] Lane, P. R., and Milesi-Ferretti, G. M.: External Capital Structure: Theory and Evidence. *International Monetary Fund Working Paper* **WP/00/152** (2000).
[24] Lane, P. R., and Milesi-Ferretti, G. M.: International Financial Integration. *International Monetary Fund Working Paper* **WP/03/86** (2003).
[25] Lane, P. R., and Milesi-Ferretti, G. M.: The external wealth of nations mark II: Revised and extended estimates of foreign assets and liabilities, 1970–2004. *Journal of International Economics* **73** (2007), 223-250.
[26] Lane, P.R., and Milesi-Ferretti, G. M.: The Drivers of Financial Globalization. *American Economic Review* **98** (2008), 327-32.
[27] Mundell, R. A.: Uncommon Arguments for Common Currencies. In: Johnson, H. G. and A. K. Swoboda (eds). *The Economics of Common Currencies*. Allen and Unwin, London, 1973.
[28] Rua, A.: Measuring comovement in the time-frequency space. *Journal of Macroeconomics* **32** (2010), 685–691.
[29] Sebnem, K. O., Papaioannou, E., and Peydro, J. L.: What Lies Beneath the Euro's Effect on Financial Integration? Currency Risk, Legal Harmonization, or Trade? *Journal of International Economics* **1** (2010), 75–88.
[30] Spiegel, M. M.: Monetary and Financial Integration in the EMU: Push or Pull? *Review of International Economics* **17** (2009), 751–776.
[31] Vahid, F., Engle, R.F., 1997: Codependent cycles. *Journal of Econometrics* **80**, (1997),199-221.

Optimal Control Model of Production-Inventory System

Pavel Pražák[1]

Abstract. Among the most important problems that are faced by manufacturing enterprises can be production planning and inventory levels. Many companies use a production-inventory system to cope with fluctuations in consumer demand for their products. If the company decides, or is forced, to store its products that are manufactured but not directly sold, additional costs for holding the non-sold products in warehouses have to be considered. The paper deals with a dynamical model of production-inventory system of a certain company. It aims to find optimal practice for the company for the given forecast time horizon. The objective function is given by a finite integral that represents the strategy of management to keep both the inventory and production as close as possible to the desired level. Since the model significantly depends on the specific function of a particular demand, which is considered here as an exogenous, models with different demand functions are studied in this paper as well.

Keywords: optimal control, Pontryagin maximum principle, production–inventory system.

JEL classification: R11, C61
AMS classification: 49J15

1 Introduction

The existence of inventory in different manufacturing enterprises is caused by time and space discrepancy between demand for the product and its availability. Contractor cannot satisfy all orders without time delay and extra costs exactly at the time when the requirement for the product arises. It can be stated that in the most production processes and in the most trades it is not possible to guarantee a fluent process without a certain level of inventory. The benefits of having inventory are that the firm have its products available to satisfy instantaneous demand and further the firm can store its production during the periods with low demand. The size of inventory and its structure is influenced not only by the ability of the management of the firm to control it but also by other factors as competitive strategy, the operational and organizational structure or costs of capital etc. Inventory can be considered as one of the most expensive and important assets of many manufacturing enterprises. High inventory incurs high holding costs and low production costs while low inventory incurs low holding costs and high production costs. In other words, on one hand, firms aim to reduce costs by reducing immediate inventory levels. On the other hand, customers become dissatisfied when inventory outages (stock outs) occur. Thus, firms struggle to find the balance between low and high inventory levels. Due to the given facts, cost minimization is the major factor in obtaining this balance. As a consequence we can summarize that good inventory control is essential for optimal amount of goods or services that are to be produced. Classical economic order quantity (EOQ) model, cf. [2] or [7], assumes a constant demand rate. It is usually solved using classical optimization tools. The more general dynamic counterparts assume a time dependent demand function. To cope with time dependent demand for the product several production-inventory systems has been developed, see [5], [9], [6], [10] or [1]. Such systems consist of a manufacturing plant and a warehouse where a finished products that are not sold immediately can be stored. As soon as the product is put into inventory the firm face to inventory holding costs of two types, see [9]. First the costs related to physical storing of the product (rent the warehouse, energy, insurance, etc.). Second the costs related to blocked money to hold inventory that restrict the ability of the firm to make an

[1]University of Hradec Kralové, Department of Informatics and Quantitative Methods, Rokitanského 61, 500 02 Hradec Králové, Czech Republic, pavel.prazak@uhk.cz

investment. In the model we assume that there is both a safety stocks that the firm wishes to hold and the most efficient level of the production to keep the efficient run of the firm, cf. [10], [9] and [1].

2 Model

We consider a manufacturing firm producing a single homogeneous product and having a warehouse for inventory. Since many firms are concerned with short and intermediate term market activities we further assume that the decision horizon that the manager faces is finite and its length is T, $T > 0$. In other words inventory of products are to be controlled over the time interval $\mathbf{I} = [0, T]$.

2.1 Dynamics of Inventory

Let $I(t)$ be the inventory level of the manufactured product at time t, $t \in \mathbf{I}$. If $D(t)$ represents the demand rate for the product at time t and $P(t)$ the production rate at the same time, then the rate of change $\dot{I}(t)$ of the inventory level $I(t)$ can be written as

$$\dot{I}(t) = P(t) - D(t),\ I(0) = I_0,\ I(T) = I_T,\ t \in \mathbf{I}, \tag{1}$$

where I_0, is the initial inventory level and I_T is the terminal inventory level. This equation means that the inventory at time t is increased by the production rate and is decreased by the demand, that can be expressed by the sales rate. We also assume that

$$I(t) \geq 0 \tag{2}$$

for all $t \in \mathbf{I}$, which means that no shortage is allowed. It is also realistic to consider that

$$P(t) \geq 0 \tag{3}$$

for all $t \in \mathbf{I}$, which means that the firm operates with a nonnegative production rate. Demand rate $D(t)$, $t \in \mathbf{I}$, is an exogenous function of the model in the present paper and has to be specified beyond the model.

2.2 Objective Function

Let $H(I(t))$ be the inventory holding cost rate and let $K(P(t))$ be the cost rate corresponding to a production rate $P(t)$ at time t. Now the objective function of the model can be written as follows

$$C(P) = \int_0^T e^{-\rho t}[H(I(t)) + K(P(t))]dt, \tag{4}$$

where ρ, $\rho \geq 0$, is a constant nonnegative discount rate. Functions H and K are assumed to be nonnegative, convex and continuously differentiable. Their specific formulas will be given later.

The firm wishes to find an optimal production rate $\widehat{P}(t)$, $t \in \mathbf{I}$, together with optimal inventory level $\widehat{I}(t)$, $t \in \mathbf{I}$, that minimize objective function (4) subject to the constraint (1), (2) and (3). This is an optimal control problem with the state variable I and the control variable P. Using Pontryagin maximum principle, see e.g. [9], [8], the necessary or sufficient conditions for optimal process $(\widehat{I}(t), \widehat{P}(t))$, $t \in \mathbf{I}$, can be found.

2.3 Optimal Process

Since the minimum of the objective function are to be find Hamiltonian function for the problem (4) subject to (1) can be written in the form

$$\mathcal{H}(t, I, P, \lambda) = -e^{-\rho t}(H(I) + K(P)) + \lambda \cdot (P - D), \tag{5}$$

where λ, $\lambda = \lambda(t)$, is the adjoint function. Recall that (2) and (3) are considered, which means that the state variable I and the control variable P are bounded. For this problem the Hamiltonian can be augmented to Lagrangian function as

$$\mathcal{L}(t, I, P, \lambda, \mu_1, \mu_2) = \mathcal{H}(t, I, P, \lambda) + \mu_1 \cdot I + \mu_2 \cdot P, \tag{6}$$

where μ_1, $\mu_1 = \mu_1(t)$, and μ_2, $\mu_2 = \mu_2(t)$, are multipliers. Now sufficient conditions for optimal process, see [8], can be stated.

Suppose that $(\widehat{I}(t), \widehat{P}(t))$ is an admissible pair in the problem (1)–(4). Suppose further that there exist functions $\lambda(t)$, $\mu_1(t)$ and $\mu_2(t)$, where $\lambda(t)$ is continuous, while $\dot{\lambda}(t)$, $\mu_1(t)$ and $\mu_2(t)$ are piecewise continuous, such that the following relations are satisfied:

$$\frac{\partial \mathcal{L}}{\partial P}\left(t, \widehat{I}(t), \widehat{P}(t), \lambda(t), \mu_1(t), \mu_2(t)\right) = -e^{-\rho t} K'\left(\widehat{P}(t)\right) + \lambda(t) + \mu_2(t) = 0, \tag{7}$$

$$\mu_1(t) \geq 0 \text{ and } \mu_1(t) = 0 \text{ if } I(t) > 0, \tag{8}$$

$$\mu_2(t) \geq 0 \text{ and } \mu_2(t) = 0 \text{ if } P(t) > 0, \tag{9}$$

and

$$\dot{\lambda}(t) = -\frac{\partial \mathcal{L}}{\partial P}\left(t, \widehat{I}(t), \widehat{P}(t), \lambda(t), \mu_1(t), \mu_2(t)\right) = e^{-\rho t} H'\left(\widehat{I}(t)\right) - \mu_1(t). \tag{10}$$

Since we consider that H and K are convex functions the Hamiltonian $\mathcal{H}(t, I, P, \lambda)$ is concave in variables (I, P). Put $h_1(t, I, P) = I$ and $h_2(t, I, P) = P$. These functions that can be used to represent the constraints are quasiconcave in variables (I, P). Then the pair $(\widehat{I}(t), \widehat{P}(t))$, $t \in \mathbf{I}$, is the optimal process in the problem (1)–(4).

In the present paper we mainly concentrate on the following situation: we consider that there is an interval $\mathbf{J} \subseteq \mathbf{I}$ such that $\widehat{I}(t) > 0$ and $\widehat{P}(t) > 0$ for all $t \in \mathbf{J}$. Using complementary slackness condition (8) and (9) we find that $\mu_1(t) = 0$ and $\mu_2(t) = 0$ for $t \in \mathbf{J}$. Now the optimal process is an interior solution to the problem (1)–(4). If we use (7) and (10) we get

$$K''\left(\widehat{P}(t)\right) \cdot \dot{\widehat{P}}(t) - \rho K'\left(\widehat{P}(t)\right) = H'\left(\widehat{I}(t)\right). \tag{11}$$

Substituting from state equation (1) into the latter equation (11) we finally find ordinary differential equation for optimal state variable $\widehat{I}(t)$ in the form

$$K''\left(\dot{\widehat{I}}(t) + D(t)\right) \cdot \ddot{\widehat{I}}(t) - \rho K'\left(\dot{\widehat{I}}(t) + D(t)\right) = H'\left(\widehat{I}(t)\right) - K''\left(\dot{\widehat{I}}(t) + D(t)\right) \cdot \dot{D}(t) \tag{12}$$

where $t \in \mathbf{J}$. For optimal control $\widehat{P}(t)$ the following modified state equation

$$\widehat{P}(t) = \dot{\widehat{I}}(t) - D(t), \tag{13}$$

is valid, where $t \in \mathbf{J}$.

2.4 Specific Costs Rate Functions

According to [4], the aim of managers of the firm is to minimize the deviations from the fixed level. The deviations can be defined as quadratic functions with known parameters inside the objective function. The given optimal control problem is then known as HMMS–type model, which is the abbreviation for Holt, Modigliani, Muth and Simon model, cf. [4]. If we adopt this idea we can better specify optimal process of the problem (1)–(4). Therefore let the quadratic functions for inventory holding costs rate and the production costs rate be

$$H(I) = \frac{1}{2}h(I - I_d)^2, \quad K(P) = \frac{1}{2}k(P - P_d)^2,$$

where h, $h > 0$, and k, $k > 0$, are positive real constants and I_d, $I_d > 0$, and P_d, $P_d > 0$, are desired constant levels of inventory and rate of production. Now according to [4] the objective function (4) can be interpreted as follows: penalties are incurred when the inventory level I and production rate P deviate from their respective desired levels I_d and P_d. In other words, see [9], the firm wants to keep its inventory level I as close as possible to the desired level I_d and similarly the firm wants to keep its production rate P as close as possible to the desired rate P_d. To summarize the quadratic terms measure the difference from desired values.

Now equation (12) can be rewritten as

$$\ddot{\widehat{I}}(t) - \rho\dot{\widehat{I}}(t) - \omega^2\widehat{I}(t) = -\omega^2 I_d - \rho\left(P_d - D(t)\right) - \dot{D}(t), \tag{14}$$

where $\omega^2 = h/k$. The characteristic equation of this linear ordinary differential equation is

$$m^2 - \rho m - \omega^2 = 0.$$

It has two distinct real roots

$$m_1 = \frac{\rho - \sqrt{\rho^2 + 4\omega^2}}{2} < 0, \quad m_2 = \frac{\rho + \sqrt{\rho^2 + 4\omega^2}}{2} > 0. \tag{15}$$

The general solution to equation (14) can be written in the following way

$$\widehat{I}(t) = a_1 e^{m_1 t} + a_2 e^{m_2 t} + Z(t), \quad t \in \mathbf{J}, \tag{16}$$

where a_1 and a_2 are real constants and $Z(t)$ is a particular solution to equation (14) that depends on the right hand side function of this equation and especially on the demand function. The particular solution can be found either by the procedure variation of constants or as a special type of function, for details see [3]. Using (1) and (16) we find the optimal production rate in the form

$$\widehat{P}(t) = a_1 m_1 e^{m_1 t} + a_2 m_2 e^{m_2 t} + \dot{Z}(t) + D(t), \quad t \in \mathbf{J}. \tag{17}$$

3 Numerical Examples

In this section we deal with particular numerical examples.

Example 1. First we consider a specific formula for demand function

$$D(t) = 0,1 \cdot t \cdot (t-6) \cdot (t-12) + 10. \tag{18}$$

A problem without discounting is assumed and particular values of exogenous variables and parameters are given, see Table 1.

exogenous variables and parameters	value
length of time horizon	$T = 12$
discount rate	$\rho = 0$
desired level of inventory level	$I_d = 10$
desired level of production	$P_d = 20$
inventory holding cost coefficient	$h = 1$
production cost coefficient	$k = 1$
initial inventory level	$I_0 = 0,01 \cdot I_d$ or $I_0 = I_d$
terminal inventory level	$I_T = I_d$

Table 1 Variables and parameters specification.

The length of time planning horizon is supposed to be 12 month (or 1 year). The specific formula of demand function can be found e.g. as a least square approximation of historical time series collected by the given firm. Here we use a cubic function with a maximal level and a minimal level of demand within the given time horizon. The initial, terminal and desired levels of inventory or the desired level of production primarily depends on decisions of managers of the firm.

Given the values of parameters the roots (15) of the characteristic equation are $m_1 = -1$ and $m_2 = 1$. Then the particular solution is $Z(t) = 0,3t^2 - 3,6t + 17,8$. The values of constants a_1 and a_2 can be found from (16) as a boundary value problem (14) with $\widehat{I}(0) = I_0$ and $\widehat{I}(T) = I_T$. For this problem the following system of linear algebraic equations can be found and solved:

$$\begin{array}{rcrcrcl} \widehat{I}(0) & = & a_1 & + & a_2 & + & Z(0) & = & I_0 \\ \widehat{I}(0) & = & a_1 e^{-T} & + & a_2 e^{T} & + & Z(T) & = & I_T. \end{array}$$

Instead of the accurate analytical solution to the given problem the graphs of the solutions are presented on the Figure 1. On the left side of this figure there is the graph of the selected demand function. In the

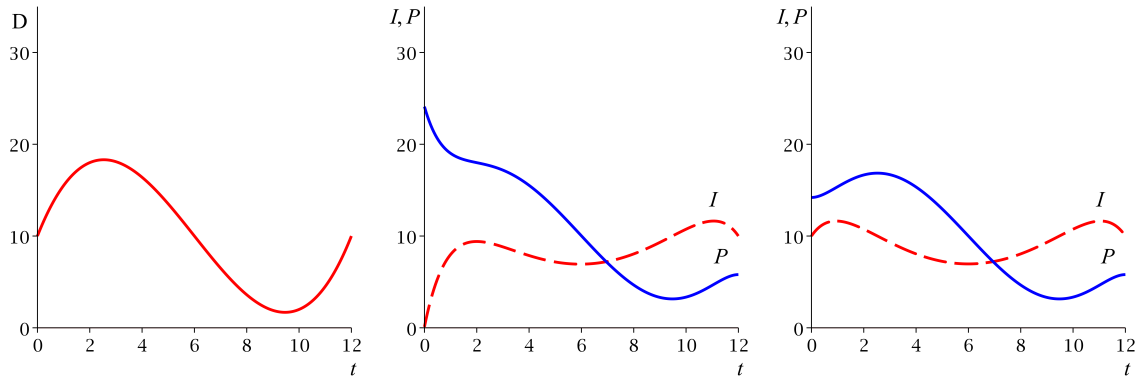


Figure 1 Demand function and the relevant optimal processes for different initial values of inventory.

middle of the figure we can find the optimal production $\widehat{P}$ as a control function and the optimal inventory level $\widehat{I}$ as its response. In this case the initial level of inventory is $I_0 = 0,01 \cdot I_d$. On the right side we can also find the optimal process, however in this case the initial level of inventory is $I_0 = I_d$. The given graphs were constructed with the help of CAS Maple.

Example 2. For the second example we consider the same values of exogenous variables and parameters as in the previous example given in Table1. For the demand function we consider a function with one maximum value in the middle of the planning horizon **I**. Particularly we consider

$$D(t) = 5 \cdot \sin\left(\frac{t \cdot \pi}{6}\right) + 10. \tag{19}$$

Instead of the accurate analytical solution we give only a graphical solution to the given problem, see Figure 2. On the left side of this figure there is the graph of the demand function. In the middle of the figure can be seen the optimal production $\widehat{P}$ and the optimal inventory level $\widehat{I}$. In this case the initial level of inventory is $I_0 = 0,01 \cdot I_d$. On the right side of the figure the optimal process can be again seen, however in this case the initial level of inventory is $I_0 = I_d$.

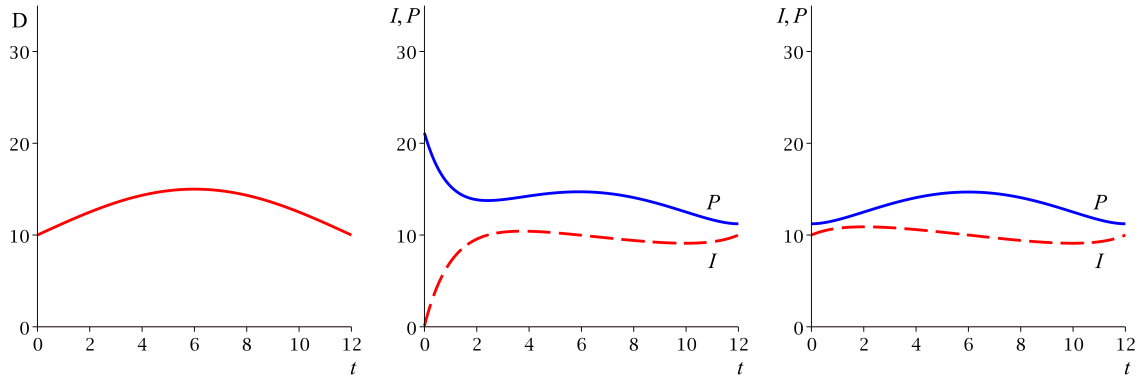


Figure 2 Demand function and the relevant optimal processes for different initial values of inventory.

The possible observations gained from the given graphical representations of optimal processes for inventory-production systems on Figure 1 and Figure 2 are:

- The closer the initial inventory level I_0 to the demanded inventory level I_d, the more similar the shape of production rate and the shape of demand function.
- If the initial level of inventory level is very low then the production rate is quite high in order to satisfy the demand for the product and to make the warehouse full with the desired level of inventory.
- The initial level of inventory has an influence on the optimal process only in the beginning of the planning period but lately it becomes insignificant.

4 Conclusion

The optimal control problem of a HMMS–type model for production–inventory system was studied in this paper. Sufficient conditions for optimal inventory levels and optimal rate of production was found by minimizing the objective function that measures the sum of penalized deviations between the state and control variables and their respective desired levels. We mainly concentrated to the interior solutions, which means that positive levels of inventory and positive values of production rate were considered. In the further research we would like to engage with more general bounds given in (2) and (3) and find a more complete discussion of possible optimal process. Other possible variations of the present model are connected with different objective functionals.

Acknowledgements

The support of the Czech Science Foundation $\#14 - 024245$ is kindly acknowledged.

References

[1] Bensoussan, A., Hurst, E., and Naslund, B.: *Management Applications of Modern Control Theory*, North-Holland Publ., New York, 1974.

[2] Balakrishnan, N., Render, B., Stair, R. M.: *Managerial Decision Modeling with Spreadsheets*, Pearson Prentice Hall, New Jersey, 2007.

[3] Coddington, E. A.: *An Introduction to Ordinary Differential Equations,* Dover Publication, Inc., New York, 1989.

[4] Holt C. C., Modigliani F., Muth J. F. and Simon H. A.: *Planning production, Inventories, and Work Force.* Prentice-Hall, Englewood Cliffs, 1960.

[5] Knowles, G.: *An Introduction to Applied Optimal Control*, Academic Press, New York, 1981.

[6] Nevima, J., Ramík, J. Application of DEA for evaluation of regional efficiency of EU regions. In: *Mathematical Methods in Economics 2010,* University of South Bohemia in České Budějovice : Faculty of Economics, 2010, 477–482.

[7] Ragsdale, C. T.: *Managerial Decision Modeling*, South-Western, Cengage Learning, 2011.

[8] Seierstad, A., Sydsæter K.: *Optimal Control Theory with Economic Application,* North-Holland, 1987.

[9] Sethi, S. P., Thompson, G. L.: *Optimal Control Theory, Applications to Management Science and Economics.* Kluwer Academic Publishers, Boston, 2003.

[10] Sprzeuzkouski, A. Y.: A problem in optimal stocks management. *Journal of Optimization Theory and Applications* **vol. 1** (1967), 232–241.

Financial threats of network stability caused by fragments tending to bankruptcy

Jaroslava Pražáková[1], Martin Pech[2]

Abstract. The companies with tendency to bankruptcy are affected not only by their own financial situation but also by financial power of their cooperative partners and their own network structure position. The paper is focused on the impact of financial situation of membership companies themselves and structure of network. The main aim of the paper is identification of network fragments with tendency to bankruptcy which could threaten the network stability.
Network structure position of the nodes was determined by the genetic search algorithm for the optimum Core/periphery bipartition.
In evaluation process are financial threats in triads find out. The process also incorporates specific Credit and Bankruptcy Prediction Models. With respect to interdisciplinary character of the described network, three models are selected: Taffler index, Kralicek Quick test, IN99 index. Triad threats level (TTL) determinates financial situation of membership companies with regard to cooperative relationships in triadic concept of the network. Based on membership to core or periphery of network, triads are divided into three groups (core, intermediate and periphery triads). Fragments with tendency to bankruptcy are composed of triads with very high threats level.

Keywords: Financial threats, Core/periphery model, credit and bankruptcy prediction models.

JEL Classification: L14, L60, G1, G3
AMS Classification: 90B10, 91G50

1 Introduction

Changing global conditions are also having an impact on future demand of almost markets. Forecasts for the year 2015 present modest growth of the traditional markets in Europe, Japan and the United States, and fast growth in the new emerging markets. The financial crisis caused the exhausting of almost all financial reserves of survived companies. For that reason, the companies are far more vulnerable to bankruptcy than years before.

The companies are affected not only by their own financial situation but also by financial power of their cooperative partners and their own network structure position. Competition always increases the risk of dissolution. The formation and dissolution of the relationships between companies receive far less attention then the literature about birth and death of companies [2]. Nevertheless, the relationships with companies threaten by bankruptcy can bring new threats. The potentially breaking links (ties) can negatively influence performance of company, part of network or network as whole. Threaten part of network has tendency to create fragments.

When any resources or activities are shared between relationships there will be either a positive or a negative connection between them. Business relationships are connected to each other. It can be illustrated by the simple example of three companies related through two business relationships [6]. No one interaction, whether it is a sale, purchase, advice, delivery or payment can be understood without reference to the relationship of which it is a part. Similarly, no one relationship can be understood without reference to the wider network [6].

It is possible to say that impact on network structure should be related to core and periphery membership. In case of companies with tendency to bankruptcy clustered in one specific part of network it should be identified as a problematic fragment. The founding of the errorless method for bankruptcy or financial health identification is in centre of the interest for many years. We refer readers for example to these credit and bankruptcy prediction models: Altman Z-Score, IN indices, Tamari index, Taffler index, Q-index (Gurčík index), Kralicek Quick test.

[1] University of South Bohemia in České Budějovice, Faculty of Economics, Department of Accounting and Finances, Studentská 13, 370 05 České Budějovice, smoloj@ef.jcu.cz
[2] University of South Bohemia in České Budějovice, Faculty of Economics, Research and development department, Studentská 13, 370 05 České Budějovice, mpechac@ef.jcu.cz

Choi and Wu [9] have stated that a dyadic framework fails to consider companies as nodes that are embedded in larger supply networks [10]. A dyad makes no reference to how a link may affect another link or a node affects multiple links; in this regard, it is not the dyad but the triad that is the fundamental building block of a network [8]. With necessity of managing bigger networks propose Choi and Wu [9] that a buyer–supplier and supplier–supplier relationship are two interdependent pieces of the same puzzle–a triadic buyer–supplier–supplier relationship.

In the foundational literature on social networks, the analysis of triads constitutes an important area of inquiry. A triad is a set of three actors and the possible ties among them. The triad is a core structure of higher-order networks [15], and since dyadic ties are embedded in triads, they represent a valuable layer of meaning for the network analyst [11]. Business triad concept is very meaningful to study how relationships among companies with financial risk and treads could lead into dissolution and breaking ties, which fragments of network are potentially endangered by bankruptcy.

Bad results of some members of network could cause death of the company and can influence other companies via relationships in network and in negative manner create above mentioned network fragments. We try to find way how to challenge problem of finding network fragments caused by bankruptcy. The methodology is demonstrated in small network case, but has potential for implementation on broader network. We try to design model of finding fragments. Paces with the most concentrated bad results companies connected together.

2 Methodology

The paper is focused on the impact of financial situation of membership companies themselves and structure of network. The research question concerns the interface between the node and the threads. The network interplays between influencing other companies and is influenced by them. Relationships provide the opportunity for the company to influence others, but the same relationships are also a force for these others to influence the company [6]. The companies with tendency to bankruptcy are affected as well as by financial power of their cooperative partners and their own network structure position. Due to this the problematic fragments could be detected. The main aim of the paper is identification of network fragments with tendency to bankruptcy which could threaten the network stability.

Core/periphery model identifies which nodes belong to the core and which belong to the periphery. Model is based on calculation the influence measure between every pair of nodes by using the model which helps to identify core and periphery parts of network [4]. The relation among the members of one group is compact, and the group can be regarded as „core“. The relation among the members of another group is loose, and the group can be regarded as „periphery“ [7]. Borgatti and Everett [4] offered a discrete formulation of core/periphery structure as a bi-partitioning problem based on the genetic search algorithm (Goldberg, 1989) implemented in software UCINET [5]. The model actually finds the optimum core/periphery bi-partition. A simple measure of how well the real structure approximates the ideal (complete connected network - Freeman star) is given by Eq. (1) together with Eq. (2).

$$\rho = \sum_{i,j} a_{ij}\delta_{ij} \tag{1}$$

$$\delta_{ij} = \begin{cases} 1 \text{ if } c_i = \text{CORE and } c_j \text{ CORE} \\ 0 \text{ if } c_i = \text{PERIPHERY and } c_j \text{ PERIPHERY} \\ . \text{ otherwise} \end{cases} \tag{2}$$

In the equations, a_{ij} indicates the presence or absence of a tie in the observed data, c_i refers to the class (core or periphery) that actor i is assigned to, and δ_{ij} (subsequently called the pattern matrix) indicates the presence or absence of a tie in the ideal image. For a fixed distribution of values, the measure achieves its maximum value when and only when A (the matrix of a_{ij}) and Δ (the matrix of δ_{ij}) are identical, which occurs when A has a perfect core / periphery structure. Thus a structure is a core / periphery structure to the extent that ρ is large [4].

The Eq. (2) is modified definition of the pattern matrix Δ where we treat off-diagonal regions of the matrix as missing data (indicates “.”). So the algorithm seeks only to maximize density in the core and minimize density in the periphery, without regard for the density of ties between these regions (recommended model). To test the robustness of the solution we run the algorithm a number of times from different starting configurations. If there is good agreement between these results then this is a sign that there is a clear split of the data into a core/periphery structure [5].

Evaluation process of financial threats starts with identification of triads. In a directed fully connected network there are $g(g-1)(g-2)/6$ triads, where g indicates number of nodes. In a directed network there are sixteen possible triad types (for an undirected network there are only 4 possible types namely 003, 102, 201 and 300[3]). For evaluation process are valuable only triads with two or all three connected nodes. The algorithm (figure 1) is run for all observed triads in network. Every triad consists of three companies: C_1 with financial threat level x_1, C_2 with financial threat level x_2 and C_3 with financial threat level x_3. The financial threat level (FTL) is value from 0 to 3 (where 0 … no threat, 1 … low threat, 2 … high threat and 3 indicates very high threat level). The process also incorporates specific Credit and Bankruptcy Prediction Models. With respect to interdisciplinary character of the network, three models are selected: Taffler index, Kralicek Quick test, IN99 index.

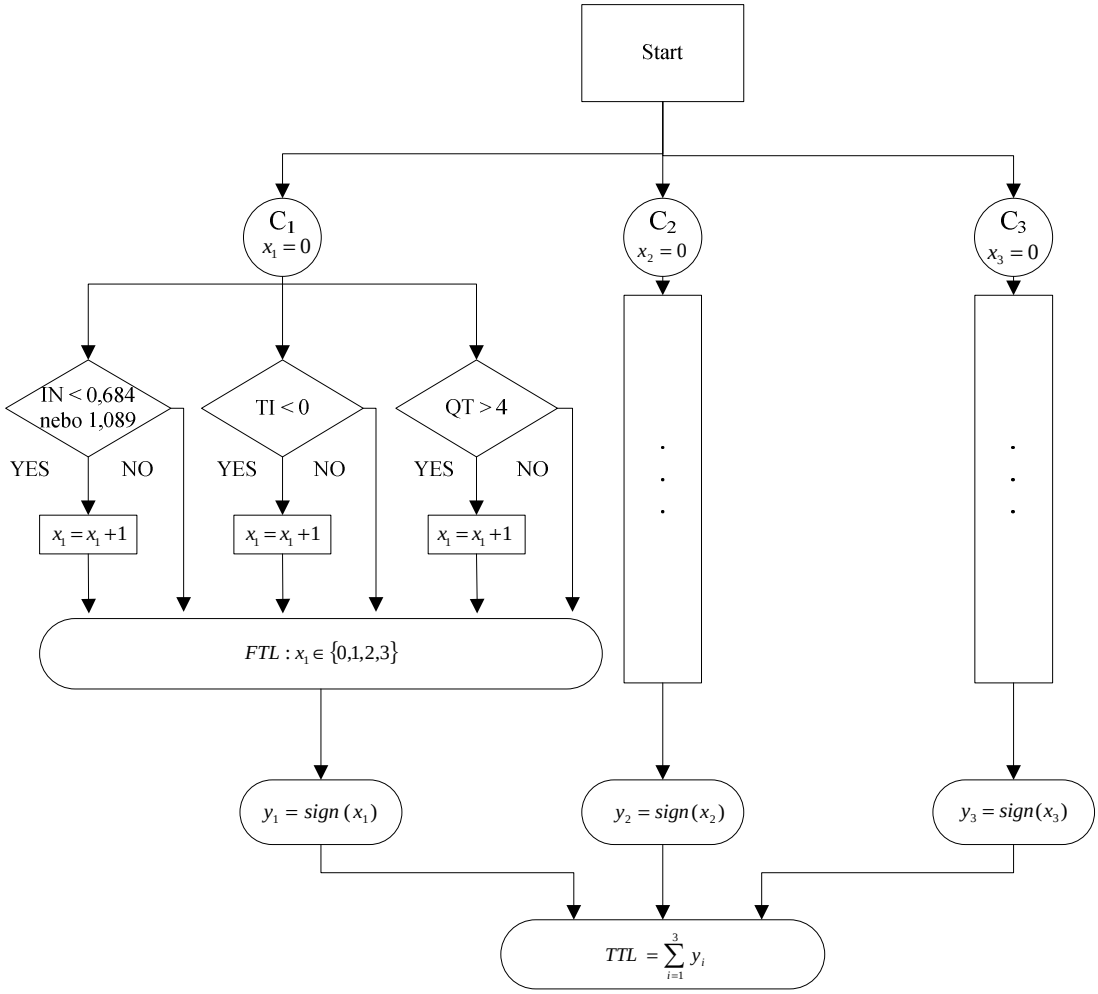


Figure 1 Evaluation process algorithm

The generic Taffler index (TI), sometimes Taffler model, is the distillation into a single measure of a number of appropriately chosen financial ratios, weighted and added. If the derived TI is above a cut-off, the firm is classified as financially healthy, if below the cut-off, it is viewed as a potential failure. Firms with computed TI below 0 are at risk of failure, and above 0 are financially solvent (more about TI algorithm and conditions for example Agarwal and Taffler [1]. Agarwal and Taffler [1] demonstrated the z-score model as true failure prediction ability and typifies a far more profitable modelling framework for banks than alternative, simpler approaches.

At the beginning, Kralicek Quick Test (QT) was widely used in German speaking countries (Germany, Austria and Switzerland). Czech scholars [13] proved that Kralicek Quick Test can be successfully applied also in East European countries. Czech scholars have developed Kralicek Quick Test indicators average values for four branches – manufacturing industry, processing industry, wholesale and retail trade.

[3]Triad types are based on triad census from UCINET software [5]. Types are labelled (abcZ) by number of reciprocated ties (“a”), unreciprocated ties (“b”), number of null ties (“c”) and letter (“Z”) used to differentiate between different triads in which are numbers “abc” same.

Index IN 99 (IN) belongs to the group of credibility models and says if company creates value for shareholders. The index is applicable on every branch of companies and therefore it is possible to use it for financial health statement in network consisted of companies with different production program (more about index IN 99 algorithm in Beranová, Basovníková and Martinovičová [3] or Neumaierová and Neumaier [12].

The result of evaluation process can be regard as calculation of triad threats level (TTL), that determinates financial situation of membership companies with regard to cooperative relationships in triadic concept of the network. TTL is calculated as a sum of normalized company financial threats levels (FTL). The financial threats levels (x_1, x_2, x_3) are normalized by signum function to normalized triad threats levels (y_1, y_2, y_3). TTL ranges value from 0 to 3 (where 0 … no threat, 1 … low threat, 2 … high threat, and 3 indicates very high threat level).

Based on membership to core or periphery of network, triads can be divided into three groups: core, intermediate and periphery triads. The core triads consist only of companies from network core. Otherwise periphery triads consist only of companies from network periphery. Intermediate triads comprise companies with different network position (at least one lie in core). Problematic fragments are composed of triads with very high threats level.

3 Results

Analysed network is composed of 19 companies (denoted by *C*+number of company) linked by supplier customer relationships. All companies in network are middle sized or big companies oriented on automotive industry and therefore very sensitive on changes caused by financial crisis. For calculating credit and bankruptcy predictions models (IN, TI, QT) are used financial results of the companies from the year 2013 when first moderate grow was indicated. Information are shared between the companies in both directions (outcoming and incoming), feedback information is always dispatched. The network structure is designed strictly based on information flow. The considered network case could be described as a directed and disconnected graph consists of 19 nodes and 21 links. Two of these links are reciprocal; control information flows from both companies are on the same level.

The network is also sparse, where the links better integrate information exchanges and are better controlled by centres. The overall network density (D) is D = 0,0673 (σ = 0,2505). It means the proportion of all possible links that are actually presented covers only about 6,73 % of potential network connections. Observed supply network suggests by its topology more chain than highly connected network [14].

Core/periphery model

In core/periphery discrete model we test CORR (starting fitness = 0,335), Density (starting fitness = 1,255) and Hamming model (starting fitness = 0,849) based on fitness measures with starting parameters: data type = positive, maximization criterion, number of iterations: 50, population size: 100. To test the robustness of the solution the algorithm we run a number of times from different starting configurations [14]. The results of the Core/Periphery Class Memberships in blocked adjacency matrix give similar results of Core/Periphery Class Membership (see Figure 2, core nodes are depicted in black colour, periphery in grey colour): Core (C013, C032, C049, C083, C106, C200) and Periphery (C016, C027, C060, C061, C076, C080, C086, C140, C143, C144, C148, C151, C165).

Financial threats evaluation process

Financial threats in triads are identified during evaluation process. For this phase only triads with two or all three connected nodes are valuable. The network has 45 triads which could be categorized according to triad census for directed graph to six groups (or 2 groups in undirected graph). Our results show:

- sub-graphs with two connected nodes (43): the out-star (021D = 17 triads), in-star (021U = 1 triad), directed line (021C = 15 triads), with mutual connections (111U = 10 triads),
- sub-graphs with all three connected nodes (2): with one mutual connection (120D = 1 triad; 120C = 1 triad).

The process incorporates 3 specific Credit and Bankruptcy Prediction Models. Table 1 shows results of TI, IN, QT and FTL values (x_i) for all companies in the network. Companies with well financial situation (TI, IN, QT) are depicted in table 1 with light grey colour and bad financial situation (TI, IN, QT) marked with dark grey. TFL values are calculated based on our algorithm (Figure 1). Higher x_i value indicates higher financial threat of company. For example company C151 has high level (x_i = 2) of financial threat (0 + 1 + 1 = 2) with regard to results of specific Credit and Bankruptcy Prediction Models. Results of financial threat in companies are used for calculation of triad threats level (TTL).

Company	TI	IN	QT	x_i
C013	0,74	1,41	1,50	**0**
C016	0,56	0,41	2,50	**1**
C027	0,43	1,18	1,75	**0**
C032	0,82	2,78	1,25	**0**
C049	0,26	0,11	2,25	**1**
C060	0,30	0,08	4,00	**2**
C061	0,39	0,59	1,75	**1**
C076	0,52	1,12	1,75	**0**
C080	0,69	1,76	2,25	**0**
C086	0,35	0,00	4,00	**2**

Company	TI	IN	QT	x_i
C106	0,50	0,95	2,00	**1**
C140	0,92	2,01	1,25	**0**
C143	1,24	3,14	1,25	**0**
C144	0,27	0,21	3,25	**1**
C148	0,67	1,56	1,75	**0**
C151	0,33	-0,59	4,00	**2**
C165	0,93	2,21	1,25	**0**
C200	1,00	1,50	2,50	**0**
C083	1,25	2,51	1,25	**0**

Table 1 Financial threats levels (TFL) of companies (authors)

Triad threats level is based on financial situation of membership companies in given triad. The normalization of entire FTL (x_1, x_2, x_3) with regard to cooperative relationships of all companies in triad is calculated by signum function. The sum of normalized values of FTL can be regard as TTL of given triad. For example selected triad consists of companies C049, C060, C061, where $x_1 = 1$, $x_2 = 2$, $x_3 = 1$. Then TFL is $sign(1) + sign(2) + sign(1) = 3$ which indicates very high level of financial threats in triad. Based on core or periphery membership of companies, we divided triads into three groups (core, intermediate and periphery triads). Table 2 shows frequency of triad types according to the network position and threats level.

Triad threats level (TTL)	**Core Triads**	**Intermediate Triads**	**Periphery Triads**	**Overall frequency**
0	3	10	2	**15**
1	4	10	2	**16**
2	1	7	2	**10**
3	0	2	2	**4**

Table 2 Triads frequency according to network position and threats level

Network fragments identification

Problematic fragments are composed of triads with very high threats level (TTL = 3). The analysed network has four triads with this feature: C151-C016-C060, C061-C060-C016, C049-C060-C016, C049-C060-C061. None of very high threaten triads is localized in core, so it refers to overall network stability. Figure 2 shows high threats level triads in periphery. Company C049 which is part of two intermediate triads with TTL = 3 is a member of the core. The results demonstrate clustering tendency of companies with bad financial situation and it creates problematic fragment which could causes instability of the network.

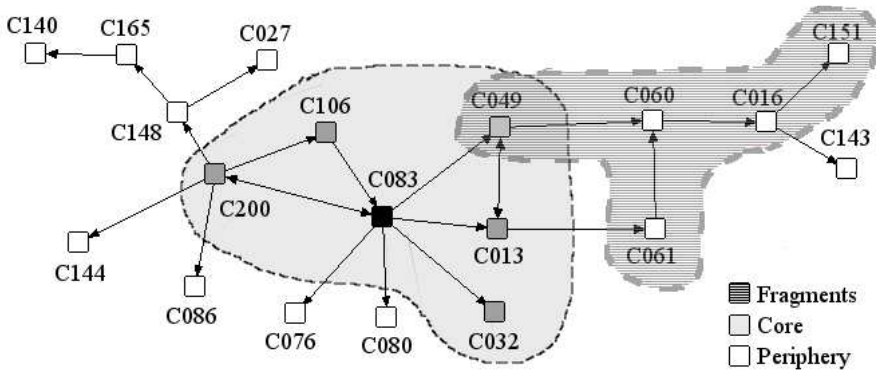


Figure 2 Core/periphery network model with network fragments (authors)

4 Conclusion

The relationships with companies threaten by bankruptcy can bring new threats especially after the financial crisis. The paper is focused on the impact of financial threats of membership companies in network interplay. Combination of triadic concept and Credit and Bankruptcy Prediction Models was used for problematic network fragments identification in evaluation process.

Based on membership of companies to core or periphery parts of network, 45 triads with two or all three connected nodes in observed network are divided into three groups (core, intermediate and periphery triads). Determined problematic fragments are composed of triads with very high treats level. The analysed network has four triads with this characteristic. None of very high threaten triads is localized in core part of the network, so it refers to overall network stability whereas high triads treats level are in periphery. The results demonstrate clustering tendency of companies with bad financial situation and create problematic fragment which could causes instability of the network.

Presented algorithm (evaluation process) was tested only on case study. Observed network has stabile structure and incorporates companies mainly with more than average financial health. We would like to test the algorithm under broader network conditions. Added to this, one question remains: Are used Credit and Bankruptcy Prediction Models the most appropriate for this purpose with respect to its number, characteristics and observed time period?

Acknowledgements

The authors thank the Ministry of Education of the Czech Republic for financial support.

References

[1] Agarwal, V., Taffler, R. J.: Twenty-five years of the Taffler z-score model: does it really have predictive ability? *Accounting and Business Research* **4** (37), 2007, 285-300. ISSN 0001-4788.

[2] Baker, W. E., Faulkner, R. R., Fisher, G. A.: Hazards of the market: The continuity and dissolution of interorganizational market relationships. *American Sociological Review* **2** (63), 1998, 147-177.

[3] Beranová, M., Basovníková, M., Martinovičová, D.: Association between values of the index IN99 and the EVA Indicator. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **2** (59), 2011, 25-33. ISSN 1211-8516.

[4] Borgatti, S. P., Everett, M. G.: Models of core/periphery structures. *Social Networks* **4** (21), 1999, 375-395. ISSN 0378-8733.

[5] Borgatti, S. P., Everett, M. G., Freeman, L. C. *UCINET 5 For Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies, 1999.

[6] Hakansson, H., Ford, D.: How should companies interact in business networks? *Journal of Business Research* **2** (55), 2002, 133-139. ISSN 0148-2963.

[7] Chang, S., Zhang, Q. Discrete Core/Periphery Structure Characters of C2 Organizations Analysis Based on SNA. In Xu, B. *2012 International Conference on Information Technology and Management Science(ICITMS 2012) Proceedings (part IV)*. Springer, Berlin - Heidelberg, 2013, 813-821.

[8] Choi, T. Y., Kim, Y.: Structural embeddedness and supplier management: a network perspective. *Journal of Supply Chain Management* **4** (44), 2008, 5-13. ISSN 1523-2409.

[9] Choi, T. Y., Wu, Z.: Triads in supply networks: Theorizing buyer-supplier-supplier relationships. *Journal of Supply Chain Management* **1** (45), 2009, 8-25. ISSN 1523-2409.

[10] Kahkonen, A. K., Virolainen, V. M.: Sources of structural power in the context of value nets. *Journal of Purchasing and Supply Management* **2** (17), 2011, 109-120. ISSN 1478-4092.

[11] Madhavan, R., Gnyawali, D. R., He, J. Y.: Two's company, three's a crowd? Triads in cooperative-competitive networks. *Academy of Management Journal* **6** (47), 2004, 918-927. ISSN 0001-4273.

[12] Neumaierová, I., Neumaier, I.: Proč se ujal index IN a nikoli pyramidový systém ukazatelů INFA. *Ekonomika a Management* **4** 2008, ISSN 1802-8470.

[13] Novotná, M., Svoboda, J.: Application of a system indices to an analysis of agricultural holdings economy in LFA regions. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **3** (58), 2010, 165-174. ISSN 1211-8516.

[14] Pražáková, J., Pech, M. The role of centres in supply network information distribution. In: *Proceedings of the 31st International Conference on Mathematical Methods in Economics 2013 - part II*. College of Polytechnics, Jihlava, 2013. 743-748. ISBN 978-80-87035-76-4.

[15] Wasserman, S., Faust, K.: *Social Network Analysis. Methods and Applications*. Cambridge University Press, London, 1995.

Delay Management in Public Transport

Vladimír Přibyl[1], Jan Černý[2]

Abstract. The paper deals with the following situation: A set of bus journeys J moves in accordance with the given timetable. Any journey j is characterized by a finite sequence of departure d and arrival a times of important segments of j. $d(i)$ and $a(i)$ mean the departure time to the i-th segment and the arrival to the end of the same segment respectively in the timetable. Moreover, a random variable $X(i)$ is given for each segment of each journey. $X(i)$ express the random running time of i-th segment. Finally a set C of quintuples (j, i, j', i', w) is given. Each quintuple corresponds with the situation, when the timetable arrival time of segment i of journey j is less then departure time of segment $i'$of journey j' and the change from journey j to journey j' is assumed at relevant segments. Attribute w expresses the upper limit of waiting time of the journey j' for the passengers changing from j. An asynchronous simulation model was created for the purpose to simulate the delay propagation in the operation of the set of journeys J. The paper summarizes the results obtained by the model.

Keywords: public transport, delay propagation, network, model, simulation

JEL Classification: C63, O18,
AMS Classification: 68U20, 90B15

1 Introduction

The paper deals with the following situation: A set of bus journeys J moves in accordance with the given timetable. Any journey $j \in J$ is characterized by a finite sequence $d_{j0}, a_{j1}, d_{j1}, \ldots, d_{jn(j)-1}, a_{jn(j)}$ where $n(j)$ represents the number of important segments of j, d_{ji-1} and a_{ji} mean the departure of it to the i^{th} segment and the arrival of it to the end of the same segment respectively in the timetable. Moreover, a random variable X_{ji} is given for each $j \in J$ and $i = 1, \ldots, n(j)$. It represents the real-word running time of the journey j through its i^{th} segment. It is supposed that all variables X_{ji}, $j \in J$ and $i = 1, \ldots, n(j)$, are mutually independent. Moreover, in accordance with [6], it is supposed that they are Beta distributed. Finally a set C of quintuples is given where $(j, i, j', i', w_{j'i'}) \in C$ means that $a_{ji} < d_{j'i'-1}$ and the change from j after i^{th} segment to the i'^{th} segment of the journey j' is assumed if the necessary waiting of the journey j' for the passengers changing from j does not exceed $w_{j'i'}$.

Remark 1. As described above, the waiting time $w_{j'i'}$ of the connecting journey j' for the delayed journey j does not depend on j. However, the model described in the sequel can be easily changed replacing $w_{j'i'}$ by $w_{jij'i'}$.

Since the real segmental running times are random, it may happen with some probability that the real arrival Ra_{ji} exceeds the timetable arrival a_{ji}, i.e. there is the delay $Da_{ji} = Ra_{ji} - a_{ji}$. Such a delay may cause further delays of the same journey j, or a delay of another journey $j' \in J$ if there exists a quintuple $(j, i, j', i', w_{j'i'}) \in C$. Such a dissemination of delay in the set J is called *delay propagation* and it is the main object the current paper is focused on.

The delay can be viewed from two perspectives. The first is the delay of vehicles or trains that serve the set of journeys J. The second consists in the delay of passengers. The model presented later enables to study both.

As one can see in [4], the passenger transport delay management theory is focused on two main issues. The first deals with "real time" delays and their management, the second studies the planning phase, i.e. the design of "robust" (with respect to delays) time tables and vehicle schedules. The last mentioned direction can be found e.g. in [1] [2] or [3]. The model described in the current paper represents a tool for displaying the delay propagation in the set J, i.e. to show what is its robustness. It represents a continuation of the paper [5]. The fundamental theoretical background is the same but the technical design and the results are quite different type. Whereas [5] has shown the delay propagation in the sequence of consecutive bus journeys, the current paper studies this process in general, rather than linear, structure of the set J.

[1] University of Economics in Prague/Faculty of Management, Department of Exact Methods, Jarošovská Street 1117/II, 377 01Jindřichův Hradec, Czech Republic, pribyl@fm.vse.cz.
[2] University of Economics in Prague/Faculty of Management, Department of Exact Methods, Jarošovská Street 1117/II, 377 01Jindřichův Hradec, Czech Republic, cerny@fm.vse.cz.

2 Model

The model starts from the situation described at the beginning of the previous chapter. First, it defines a stratification of all segments of all journeys from the set J into the *layers* L_0, L_1, … as follows:

$L_0 = \{(j', i'): j' \in J, i' \in \{1, \ldots, n(j)\}$, there exist no (j, i) such that $(j, i, j', i', w_{j'i'}) \in C\}$

$L_1 = \{(j', i') \notin L_0 : j' \in J, i' \in \{1, \ldots, n(j)\}, [(j, i, j', i', w_{j'i'}) \in C] \Rightarrow [(j, i) \in L_0]\}$

$L_2 = \{(j', i') \notin L_0 \cup L_1 : j' \in J, i' \in \{1, \ldots, n(j)\}, [(j, i, j', i', w_{j'i'}) \in C] \Rightarrow [(j, i) \in L_0 \cup L_1]\}$

…

$L_{k+1} = \{(j', i') \notin L_0 \cup L_1 \cup \ldots \cup L_k : j' \in J, i' \in \{1, \ldots, n(j)\}, [(j, i, j', i', w_{j'i'}) \in C] \Rightarrow$

$\Rightarrow [(j, i) \in L_0 \cup L_1 \cup \ldots \cup L_k]\}$

…

2.1 Simulation without a "Breakdown"

A "breakdown" means an unexpected, unlikely event, which causes a delay of δ minutes of d_{j0} for some $j \in J$. Now, it is supposed that no breakdown exists and the real departures $Rd_{j0} = d_{j0}$ for all $j \in J$. The **first course of the simulation** consist of s steps where L_s is the last nonempty layer, i.e. $L_s \neq \varnothing$ and $L_i = \varnothing$ for each $i > s$.

Initial step: Put $Rd_{j0} = d_{j0}$ for all $j \in J$. Find the value of X_{j1} and put $Ra_{j1} = Rd_{j0} + X_{j1}$ for each $j \in J$

Recursive kth step: for $k = 1, \ldots, s$: For each $(j', i') \in L_k$ find $X_{j'i'}$ and put:

$Rd_{j'i'} = \min\{(d_{j'i'} + w_{j'i'}), \max\{d_{j'i'}, \{Ra_{ji}: (j, i, j', i', w_{j'i'}) \in C\}\}\}$ and $Ra_{j'i'} = Rd_{j'i'} + X_{j'i'}$

The first course of the simulation can be repeated many times, e.g. 10,000 and the estimations of mean arrivals and mean delays of all journeys can be calculated.

Remark 2. This model is primarily devoted to the simulation of vehicle delays. If it could be used for the calculation of passenger delays, it would be necessary to add a computing procedure finding the next suitable journey, if the passenger misses the connecting one.

2.2 Simulation with a "Breakdown"

It is supposed that $Rd_{j''0} = d_{j''0} + \delta$ for some $j'' \in J$ and $Rd_{j0} = d_{j0}$ for each $j \in J - \{j''\}$. Then the only change with respect to the subchapter 2.1 is that the sentence "Put $Rd_{j0} = d_{j0}$ for all $j \in J$." in the initial step is replaced by "Put $Rd_{j''0} = d_{j''0} + \delta$ and $Rd_{j0} = d_{j0}$ for each $j \in J - \{j''\}$".

2.3 Implementation

The above described theoretical model was implemented by means of Visual Basic for Application environment in MS Access. Despite the fact, that there probably exist different specialized simulation environments, we have chosen the environment of VBA. The main reason is authors' familiarity with this environment. There was no need to study any new environment and its programing language. In addition the MS Access environment brings easy way to input and output data and to develop background for testing automation. Moreover the performance of simulation model implemented in this environment is acceptable. For example 10,000 simulation courses with the 54 journeys (each with 4 segments) take about 5 minutes including writing the results to output tables.

By means of implemented simulation model we can observe for each segment of each journey in the set J these parameters:

- mean delay at arrival to the end of segment,
- mean delay at departure from the beginning of segment,
- histogram of delays at arrival to the end of segment,
- histogram of delays at departure from the beginning of segment,
- probability of delay at arrival to the end of segment,
- probability of delay at departure from the beginning of segment,
- probability, that passengers miss the connecting journey.

Input data are the sets *J* and *C*. Some examples of results obtained are presented in next chapter.

3 Results

3.1 Partial verification of model

The first tests with the simulation model we carried out to verify the model functionality in cases where it was possible due to some reference results or easy theoretical computation.

Simulating the same simple linear structure of segments without changes with the same parameters of distribution of random variable *X* representing the real running time of the segments as in [5] has led to the same results as we have presented in [5]. Of course with an accuracy depending on the number of simulation courses.

The functionality with possible change was verified at elementary structure of 2 journeys. The structure is shown on **Figure 1**. Journey 1 consist of 2 segments 1.1 and 1.2. Journey 2 has only one segment 2.1. Change is assumed from segment 2.1 to 1.2 (depicted by rounded arrow). Let's assume that the segments 2.1 and 1.1 have the same parameters and the same probability of delay. The number p then expresses the probability of delay at arrival to the end of segment 2.1 and 1.1. The number p' expresses the probability of delay at departure at the beginning of segment 1.2. It is obvious, that in case of no possible change or $w_{1,2} = 0$ the probability $p' = p$ and in case of possible change with $w_{1,2} = \infty$ the probability $p' = p+(1-p)p$.

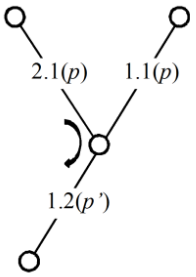


Figure 1 Elementary testing structure of set *J*

We have simulated this structure several times with different parameters and in all cases the results match the above described theoretical assumption.

3.2 Main simulation

As we already mentioned above the main object the current paper is focused on is to study delay propagation in general structure of the set *J*. In order to meet this goal we have performed a number of simulations on general mesh type structure of journeys. The structure is depicted on **Figure 2**.

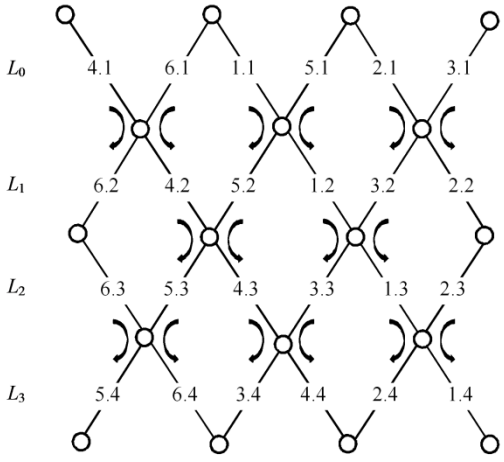


Figure 2 General testing structure

It consists from 6 journeys and each journey consists from 4 segments, which are stratified to layers L_0, L_1, L_2 and L_3. The labels of segments are in the form *j*.*i* where *j* is the number of journey and *i* is the number of segment. Rounded arrows depict the assumed changes.

Model settings

We have used simplified settings of model parameters for all performed simulation cases. The parameters (a, b, α, β) of Beta distribution of random variables X were set to: $a = 20$, $b = 24$, $\alpha = 2$ and $\beta = 4$ for all segments. It means that real running times in minutes $x_{ji} \in (20;24)$ for each $j \in J$ and $i = 1, \ldots, n(j)$. For these parameters, the mean value of all variables X was equal to 21:20 [min:sec]. Moreover we assumed, that there are no waiting times in the stations so $a_{ji} = d_{ji}$ for each $j \in J$ and $i = 1, \ldots, n(j)$. Let us denote t_{ji} the timetable running time of the i^{th} segment of j^{th} journey. All the timetable running times t_{ji} were set to the same value t for each particular simulation case. Finally all the waiting times $w_{j'i'}$ were set to the same value w for all the quintuples (j, i, j', i', $w_{j'i'}$) $\in C$ for each particular simulation case.

We have performed 54 simulation cases with all possible combinations of parameters t and w settings for both situations without "breakdown" and with "breakdown". **Table 1** brings the overview of used values.

t [min:sec]	***w*** [min:sec]
20:00	0:00
20:30	1:00
21:00	2:00
21:30	3:00
22:00	4:00
22:30	unlimited
23:00	
23:30	
24:00	

Table 1 Values of parameters t and w used in simulation cases

Simulation without a "Breakdown"

This section shows some examples of result obtained in situation when $Rd_{j0} = d_{j0}$ for all $j \in J$. **Figure 3** depicts the progression of mean delay at arrival to the segment end of journey 1 for different values of maximal waiting time for changes w. The mean delay increases with increasing value of waiting time w. This fact corresponds to the theoretical assumption.

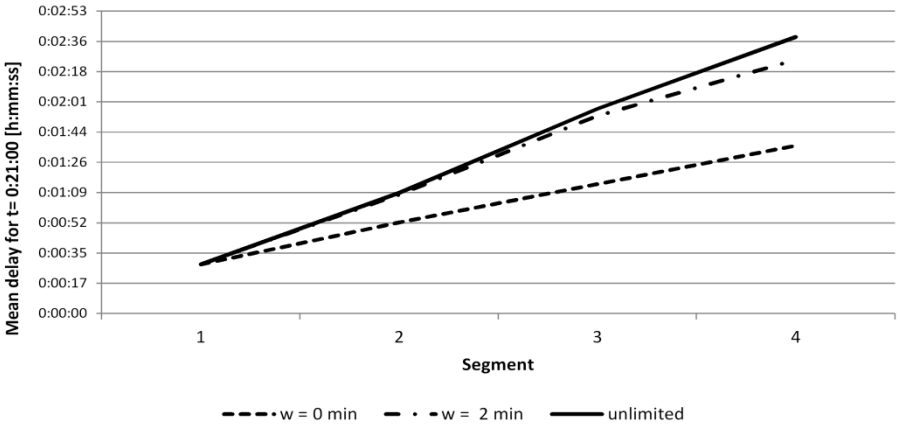


Figure 3 Mean delay at arrival to the segment ends of journey 1

Figure 4 brings more detail description of delay at arrival to the end of 4th segment end of journey 1 in form of histogram. One can see the growth of delay which leads to increase of mean delay observable in the graph on **Figure 3**.

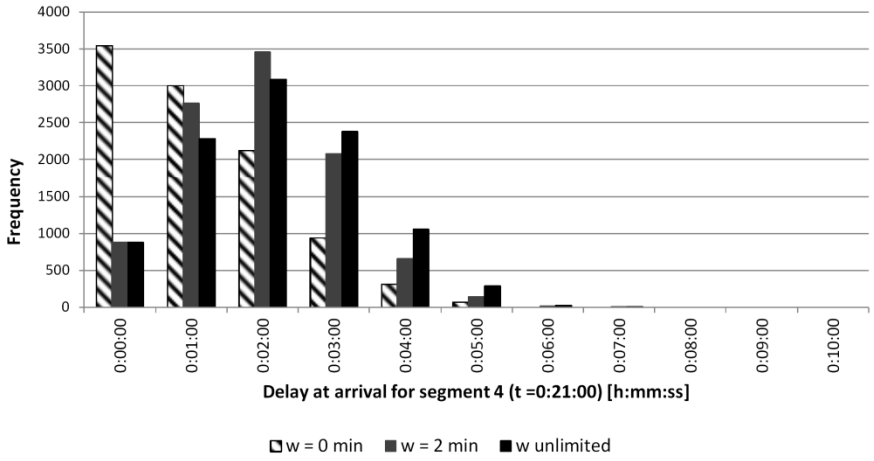


Figure 4 Histogram of delay at arrival to the 4th segment end of journey 1

Simulation with a "Breakdown"

This section shows the example of result obtained in situation when $Rd_{j''0} = d_{j''0} + \delta$ for $j'' = 5$ and $Rd_{j0} = d_{j0}$ for each $j \in J - \{j''\}$. An example of propagation **Figure 5** shows the propagation of $\delta = 10$ minutes delay at the departure of 1st segment of journey 5 in case of t = 21:30 [min:sec] for unlimited waiting time w and for $w = 2$ minutes. Journey 5 with breakdown delay is emphasized. The upper numbers in labels of segments express the mean delay at the arrival to the end of segment in situation without breakdown. The lower ones express the mean delay at the arrival to the end of segment in situation with breakdown. The mean delays influenced by breakdown delay of journey 5 are emphasized. The results correspond to the theoretical assumption that limited waiting time w reduces the delay propagation.

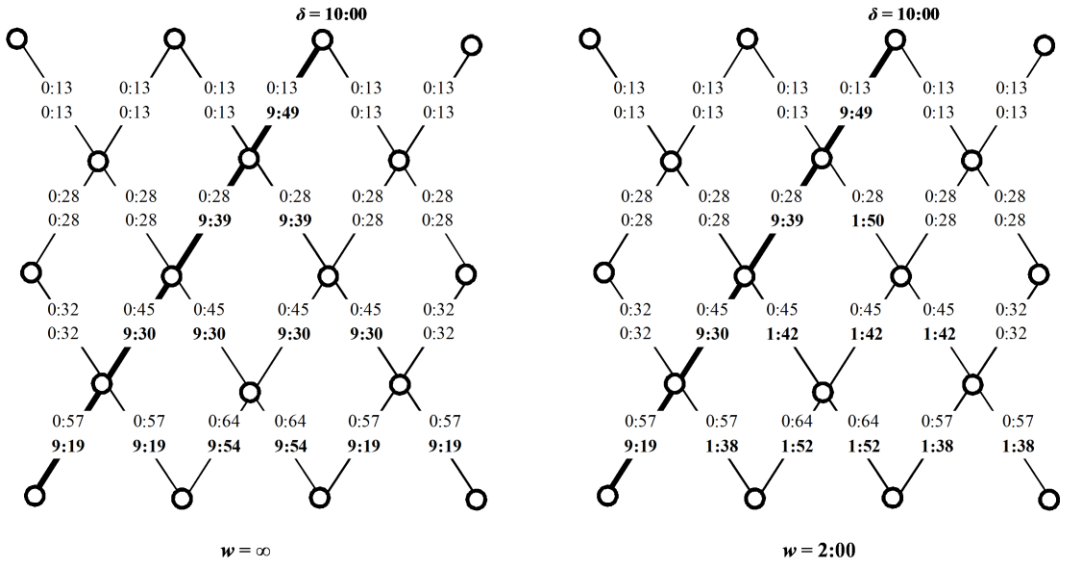


Figure 5 Propagation of "breakdown" delay at the departure of journey 5

4 Conclusion

First, the practical situation to be modelled was described. Its fundamental structure consists of

- the set J of bus journeys, characterized by their timetable and by random variables representing their factual segmental running times,
- the set of mutual linkups enabling the transfer of passengers from bus to bus.

Then the asynchronous simulation model was created for the purpose to simulate the delay propagation in the operation of the set of journeys *J*.

Finally, the paper summarizes the results obtained by the model. One of them is the fact that model yields the same results as [5] in modelling the delay propagation in particular "unbranched" structure of the set *J*.

Expected future research may be focused on the question, whether it is possible to find a model enabling a numerical solution of the same problems, solved by simulation techniques in the current paper.

Acknowledgements

The authors would like to thank the Czech Science Foundation for the support of the project 402/12/2147 whose results are described in this paper.

References

[1] Huisman, D., Freling, R. and Wagelmans, A.: A Robust Solution Approach to the Dynamic Vehicle Scheduling Problem. *Transportation Science* **38** (2004), 447-458.
[2] Kramkowski, S., Kliewer, N. and Meier, C.: Heuristic Methods for Increasing Delay-Tolerance of Vehicle Schedules in Public Bus Transport. *Proc. of MIC 2009*, Hamburg, Germany, July 13-16, 2009.
[3] Naumann, M., Suhl, L. and Kramkowski, S.: A Stochastic Programming Approach for Robust Vehicle Scheduling in Public Bus Transport. *Procedia-Social and Behavioral Sciences* Volume: **20**, 826-835
[4] Schöbel, A: *Optimization in Public Transportation: Stop Location, Delay Management and Tariff Zone Design in a Public Transportation Network*. Springer Optimization and Its Applications, Volume 3, 2006
[5] Přibyl, V. and Černý, J.: A Problem of Delay Management in Transport. *Proc. of the 31st International Conference on Mathematical Methods in Economics 2013*, Jihlava, Czech Republic, 11-13 September 2013, pp. 755-760.
[6] Stevenson, W.J.: *Introduction to Management Science*. Irwin, Boston 1989.
[7] Vašíček, R. and Černý, J.: The GOP-1 method and its use in the time-table preparation. *Rail International* **8** (1977), 98- 103.

Inference about product reliability by the analysis of complaints as a strategy for manufacturing process optimization

Angelina Rajda-Tasior[1]

Abstract. According to the strategy of many companies and also one of the fundamental areas is to improve the quality. Look on the quality from the perspective of complaints analysis is not only help to identify the most frequently occurring faults and shortcomings, but also allows to find the way to optimize processes which is the key to the development and optimization to identify and eliminate losses conceived in different dimensions. Special feature of all the characteristics which determine the quality of the product is its reliability.

In this article author is looking for the characteristics of the product reliability in the context of the complaints. As product reliability is the probability that the system or component successfully works at a specified time without a breakdown, taking into account the characteristics of the random changes of the object author tried to specify the probability of damage, the probability of failure-free operation and the probability of failure intensity for products.

The selected approach will minimize the cost of research as it was a non-destructive research. The study was conducted on a large database of complaints which applicable furniture.

Keywords: product reliability, life time of product, exponential distribution, quality, complaints analysis, failure

1 Product reliability

Beginning of a scientific approach to study the reliability of products is attributable to second part of 1950s of the XX century. It is widely considered that work that inspired many scientists to seek methods for testing the reliability is thesis of J. von Neumann indicating the possibility of constructing a reliable products of unreliable parts. Each production process is exposed to a large amount of problems. They are often associated with breakdowns and quality defects. As writes M. Imai [1] deviations from the norm is normal situation as always occurs only two possible situations: the process is stable or unstable. Therefore, in order to identify possible difficulties in the practice there are used some techniques which are typically included in the solutions of quality management.

In a practical approach, statistical methods in quality management is primarily statistical process control (SPC), Measurement Systems Analysis (MSA), statistical receiving plans and statistical methods to improve the process (e.g. DOE). Also widely used statistics and probability theory takes place in the analysis and evaluation of reliability. On one side the growing demands of users, make that demands on the reliability of products are growing fast and on the other side, the producers have difficulties in control the reliability and quality which goes together. Traditional methods of quality control which consist control of elements in product and readiness testing products are not sufficient in many cases. This raises the need to seek new methods which can determine the degree of reliance in the product at predetermined conditions.

An innovative approach in reliability analysis can be research information about products as an alternative to research of products. Only information about defects of products and time are required which are submitted in the form of customer complaints. Comprehensive database with information gathered can be the basis for inference about the reliability of products. The fundamental measure of product reliability is time of correct work it means the time from the start until the first failure. Failure means lack, defect, flaw, shortcoming, weakness etc. All that information will need to be recorded in the database and can be used as statistical material in that work.

[1] University of Economics in Katowice, Statistics department, 1 Maja 50, Katowice, angelina.rajda@gmail.com

1.1 Quality and product reliability

Quality in the context of reliability is defined as the set of attributes determining the degree of usefulness of product in specific conditions of exploitation and in accordance with its intended purpose. From the user point of view in set of characteristics which determine the quality of the product can be specified some features which are presented in table 1. Because the reliability of the product is a statistical or probabilistic feature so the quality of the product, when the reliability is one of the crucial characteristics of quality, it is a statistical or probabilistic feature as well. Therefore reliability make difficulties to quantify a quality.

Subsets of features that determine the quality of the product	**Characteristics of subsets**
Technical characteristics	construction (geometric dimensions, weight, surface condition, physico-chemical characteristics, technical and operating parameters), reliable performance, product functionality, performance, efficiency
Functional characteristics	usefulness, convenience and safety of use, reliability, durability, repairability
Aesthetic characteristics	external appearance, accuracy of execution and shapes, proportions, color and graphic solutions, the degree of compliance with the requirements, aesthetics of packaging
Economic characteristics	purchase costs, depreciation, exploitation
Additional characteristics	time and cost of production preparation, the amount of production equipment, size of the area of production, labor intensive, the number of types of semifinished products and raw materials, the amount of materials etc

Table 1. Quality characteristics in reliability theory [2]

It should be noted that normally in general opinion the price and the cost of manufacture the product is not included in the characteristics of product quality. Because when the price of the article will increase and the product does not change, user will not feel a change in the quality point of view. Comparably it is with the cost of manufacture. If the cost of producing in one factory is higher than in the other, and the product of both factories are the same, than it is deemed to that the quality of the products is the same [3]. Thus, among all the characteristics which determine the quality of the product characteristic feature there is its reliability.

Because product reliability is a statistical or probabilistic feature, meaning that for estimating the reliability of the product which we research, we need to make observation of the lots of the product within a certain time of its exploitation.

2 Measures of product reliability

Product reliability from specialized point of view, is the continuity in time certain measures usefulness and the ability to perform specific tasks. It is a characteristic formally defined as the probability that the product will function properly under certain exploitation conditions and in certain period of time. In the process of exploitation under the influence of the impact of external and internal factors e.g. temperature, load, etc. individual components of the product undergo changes**.** Then the individual parts of the product may exceed of acceptable range of specified requirements. In this way there arise malfunctions and damages. Because those factors are inherently random processes and the moment of occurrence the non-compliance is a random variable, and also life time of product in each steps will be a random variable. In that case it must be accepted that the measure of product reliability will be determined by characteristics of the random variable.

2.1 Elements of product reliability

Reliability as writes W. Wieromiejczyk [4] contains four components as result from the features of each product: failure-free, susceptibility for repair, resistance to storage and durability.

Definition 1. failure-free *(function correctly) is a characteristic of the product to maintain the ability to work at a certain time and under certain conditions without any stoppages.*

Definition 2. Susceptibility for repair *means that the product is adapted for repair.*

Definition 3. Resistance to storage *is a feature of product to maintain certain characteristics after storage and transport.*

Definition 4. Durability. *The characteristic of product, which means that the product is able to work until the limit condition, which is defined as a condition in which the further use of the product is not possible.*

Each component is associated with the random variable which has a time dimension. For the characteristic called as reliability specified attribute is the time without breakdown - T. For the susceptibility on repair specified attribute is the time which is needed for repair - T_n. For resistance to storage specified attribute is the time when product will loss quality - T_p. For the durability specified attribute is the time to occurrence of a specific limit condition - T_g.

2.2 Elements of product reliability

Distributions of random variables T, T_n, T_p, T_g in any possible form as functions of time in the interval $[0, +\infty]$ are called the reliability characteristics of the product. In reliability theory is distinguished by some characteristics A – E:

A. *Unreliability function* (cumulative distribution function) the probability that product will not perform its intended function for a given interval of time under specified operating conditions. F(t) is usually called the unreliability at time t. It represents the probability of failure in the interval [0, t]. The probability of failure in the interval (t_1,t_2] equals F(1) - F(2)

$$F(t) = P(T \le t) \tag{1}$$

B. *Reliability (survival) function* - probability of surviving at least till age t. P (T > t) means the probability that the item is still working at time t.

$$R(t) = P(T > t) = 1 - F(t) \tag{2}$$

C. *Probability density function, life time density* – this function describes the failure behavior of system over time.

$$f(t) = \frac{dF(t)}{dt} \tag{3}$$

D. *The failure rate function, hazard function* (usually represented by the Greek letter λ) is very useful quantity. This is defined as the probability of component failing in one (small) unit of time. It characterizes a risk of damage.

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{4}$$

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{-R'(t)}{R(t)} = -\frac{d}{t}\ln R(t) \tag{5}$$

2.3 Product reliability by using the exponential distribution

Estimating the reliability function using statistical methods requires a large number of experiments. These experiments need to be done always for a certain, limited period of time because that information about the reliability of the product refers only to that specified time. Basically we know nothing about the reliability of the element outside of that time means that the reliability function cannot be extrapolated. Because it is not enough statistical information about the distribution of damage it is natural to use the exponential distribution which has one parameter of the distribution which is positive. Additionally, in a situation when the real function of reliability is not exponential that the exponential function gives quite a good approximation of searched function. Exponential distribution is the basis of research plans recommended in the standard PN-77/N-04021 [4]. It is one of the most commonly used distributions in reliability and it is used to predict the probability of survival to a particular time. If λ is the failure rate and t is the time, then the reliability R can be determined by the following equation: ($t \ge 0$; $\lambda \ge 0$)

$$R(t) = e^{-\lambda t} \tag{6}$$

from the above we obtain:

$$F(t) = 1 - e^{-\lambda t} \tag{7}$$

$$f(t) = \lambda e^{-\lambda t} \tag{8}$$

$$\lambda(t) = \frac{f(t)}{R(t)} = \lambda \tag{9}$$

thus, the expected time for working product with reliability R is:

$$T = E(\tau) = \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda} \tag{10}$$

therefore, the reliability function is often written as:

$$R(t) = \exp(-\frac{t}{T}) \tag{11}$$

The hypothesis about an exponential distribution can also be checked statistically [5].

3 Case study

Reliability analysis of products includes mainly the analysis of lifecycle product. The aim in that work is to identify the reliability of the product and to get information about quality of testing products. It is an attempt of a new way to monitoring quality while minimizing costs and interference in production and in new products. This analysis is based on data, which are extremely valuable information, when they are collected.

For this purpose as data series was a complaints from the database of the furniture company. Data refer to entries from one month. The object of research is 203 products that consumers reported as the complaints in July 2013. As figure 1 shows reason of complains and their percentage in all. Products are simple, but functional and belong to the class of objects which can be fixable. Therefore, from the economics point of view, these tests can be regarded as indestructible and without interfering in the production process, which is positive. Additionally if the useful life is long it may be not practical to wait until products or components fail. It also may be too expensive to test large samples [6]. Generally, for estimating the reliability of the product it is obligate to have data with the observations of the lots of the product within a certain time of his life. The aim of the study is to identify the reliability of the product and as a result will be the assessment of the quality of products. Research consists in checking the product durability and exactly the life of the product from the time $t = 0$ when the product started to work (purchase of the product by the consumer) to time $t = T$, when damage has occurred in the product (notification of complaint about defective product, which is registered in database).

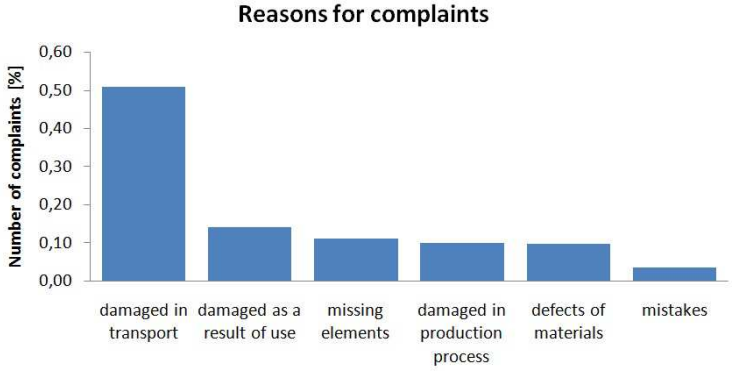


Figure 1. Reasons for complaints reported by client to furniture company

3.1 A simplified model of stability probability of failure

The first stage of the study is a simplified model of stability probability of failure (fixed risk). Set of 203 data considered as a set of 203 items (products). According to the expression: *5 * LOG10 (203)* created 12 intervals with a span of 64 days. Figure 2 shows the time when clients reported damage (failure). It presents the quantification of the complaint products. It shows the lifecycle of products from date of purchase of the product by the consumer until the notification of its damage. From the graph you can read how many products were damaged at a particular time from purchase. Then constructed histogram of 203 data points with an exponential

density curve overlaid which is on figure 3 and shows the lifecycle of products in two-month intervals. In this model, the probability of failure is constant the same for the first or the next day. This model takes into account only the random damage independent of the actual state of the product. The curve plotted on the graph is matched to the histogram of the exponential distribution. Data are reasonably close to the fitted line and it can be conclude that the exponential distribution is an appropriate choice.

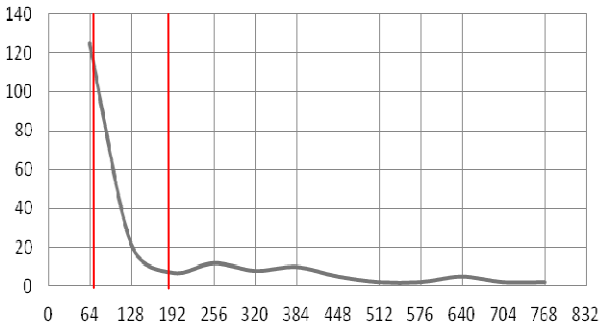


Figure 2. Quantity of complaints in the intervals of 2 months

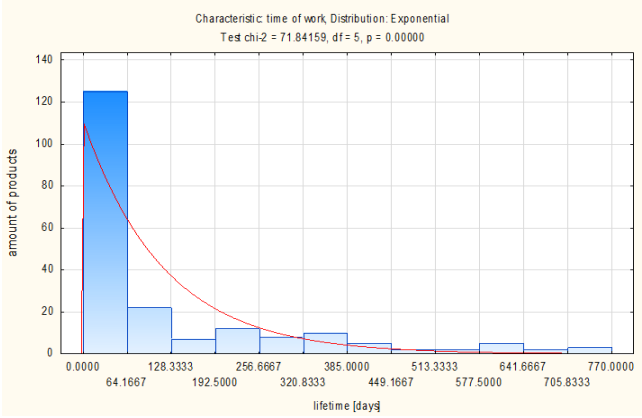


Figure 3. Histogram with an exponential density curve overlaid

3.2 The hazard curve

In reality, however, the probability of failure of the product varies and depends on many factors. Therefore, the probability of failure is higher for the new and old product. Statistics show two additional factors, which are also random. The first case concerns defects which are not found by the quality controllers (in checking process) and as the result of wrong transport and storage. The second factor is the consequence of normal wear / aging of components of the product. Due to the shape of the intensity damage curve (figure 2) the axis of the time can be divided into three intervals calculated as (0; 64), (64; 700) and (700; ∞).

First interval of intensities has elevated values. There are damaged products in less than 2 months life in the set of 203 products. Reason of these defects can be: damages during bringing, transportation or setting of the product and also condition of products whose functionality were defective within several days of use. From that reason the intensity in the interval (0,30) is very large. In the second interval (30; 700) which covers the entire warranty period until two years from the purchase is called a normal life and damage intensity is approximately constant. The last interval, which was not marked on the chart, is called post warranty period. This is the time when consumers report complaints called as payable or the period of the last few months before warranty end. In this work option 1 should be taken, as a rule, that the products in that company are repaired rather than replaced with new ones. Figures 4 and 5 presents the reliability function and the function of risk. It includes two additional factors by using the Weibull probability distribution which is a generalization of the exponential distribution.

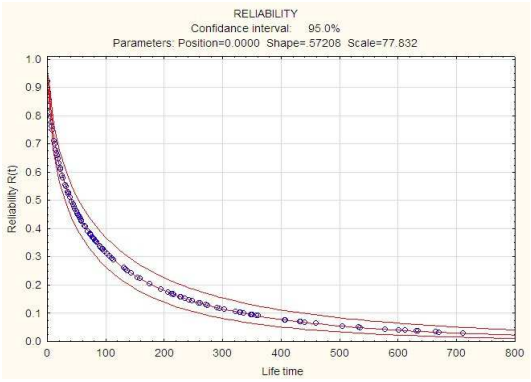


Figure 4. Summary of Weibull analysis in Statistica v.10. Product reliability

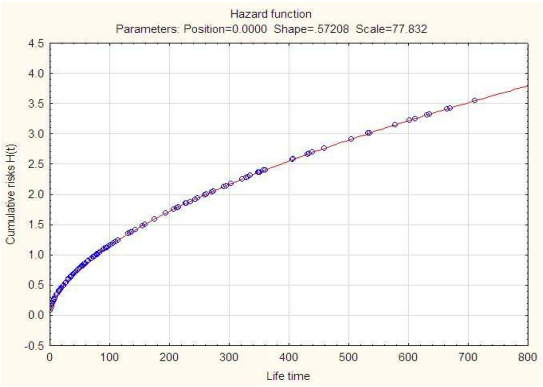


Figure 5. Summary of Weibull analysis in Statistica v.10. Hazard function

Conclusions

Reliability of product has gained increasing importance in the last few years in manufacture organizations. Reliability of products mainly refers to lifecycle of product. It seemingly means that the analysis cannot be empirically proven without products testing, but an innovative approach in that work is to show lifecycle of products with no costs, no testing, no loss time. This study is concentrated on all information from customer complaints which means that there was no destructive testing. The beginning theory was that comprehensive database with information gathered can be the basis for inference about the reliability of products. This is very optimization way of identify the most frequently occurring fault. It can be also way to eliminate losses conceived in different dimensions. As one of the fundamental areas is quality with reliability together, because "product has low quality without long reliability". Improved reliability through continuous improvement in product design and manufacturing techniques is far more effective in reducing quality problems [7]. This work is just preliminary study on reliability because the essential reason is to elicit a maximum information about the damaged product in order to prevent a similar situation in the future.

References

[1] Imai M.: *Gemba Kaizen: Zdroworozsądkowe, nisko-kosztowe podejście do zarządzania* [Common-sense, low-cost approach to management], Wydawnictwo MT Biznes, Warszawa, 2006

[2] Wieromiejczyk W.: *Jakość i niezawodność wyrobów. Studium z zakresie podwyższania trwałości elementów maszyn* [The quality and reliability of products. The study to increase the durability of machine elements], Warszawa, 1977

[3] Kiliński A.: *Jakość* [Quality], Wydawnictwo Naukowo-Techniczne, Warszawa, 1979

[4] PN-77/N-04021 *Niezawodność w technice. Plany badania w przypadku rozkładu wykładniczego* [Reliability in the technics. Investigation plans for the exponential distribution]

[5] Sołojew A. D.: *Analityczne metody w teorii niezawodności* [Analytical methods in the theory of reliability], Wydawnictwo Naukowo-Techniczne, Warszawa, 1983

[6] Bobrowski D., Ciszak J., Fidelis E., Kosmowska A., Smalko Z.: *Jakość i niezawodność maszyn i urządzeń* [Quality and reliability of machines and equipment], Stowarzyszenie Inżynierów i Techników Mechaników Polskich, Warszawa, 1979

[7] Meeker W. Q and Escobar L.: *Statistical Methods for Reliability Data*, John Wiley & Sons, 1998.

Pairwise Comparison Matrix With Fuzzy Elements

Jaroslav Ramík [1]

Abstract. This paper deals with pairwise comparison matrices with fuzzy elements from abelian linearly ordered group (alo-group) over a real interval. We generalize the concept of reciprocity and consistency of pairwise comparison matrices with triangular fuzzy numbers (PCFN matrices). We also define the concept of priority vector which is a generalization of the crisp concept. Such an approach allows for a generalization dealing both with the PCFN matrices on the additive, multiplicative and also fuzzy alo-groups. Numerical examples are presented to illustrate the concepts and derived properties.

Keywords: multi-criteria optimization, pair-wise comparison matrix, fuzzy elements

JEL classification: C44
AMS classification: 90C15

1 Introduction

Fuzzy elements of the pairwise comparison matrix (PCF matrix) can be applied whenever the decision maker (DM) is not sure about the preference degree of his/her evaluation of elements in question.

A *decision making problem (DM problem)* can be formulated as follows:
Let $X = \{x_1, x_2, ..., x_n\}$ be a finite set of alternatives ($n > 2$). The aim is to rank the alternatives from the best to the worst (or, vice versa), using the information given by the DM in the form of an $n \times n$ PCF matrix.

The DM can acknowledge fuzzy pairwise preference data as imprecise knowledge about regular preference information. The fuzzy interval preference matrix is then seen as constraining an ill-known precise consistent comparison matrix. An ordinal *ranking* of alternatives is required to obtain the best alternative(s), however, it often occurs that the DM is not satisfied with the ordinal ranking among alternatives and a cardinal ranking i.e. *rating* is required.

In the recent literature we can find papers dealing with applications of pairwise comparison method with fuzzy quantities, when evaluating e.g. regional projects, web pages, e-commerce proposals etc., see [8], [3]. The earliest work in Analytic Hierarchy Process (AHP) using fuzzy quantities as data was published by van Laarhoven and Pedrycz [7]. The issue of consistency in AHP using fuzzy sets as elements of the matrix was first tackled by Salo in [6]. Later, Ramík and Korviny in [5] investigated inconsistency of pairwise comparison matrix with fuzzy elements based on geometric mean. They proposed an inconsistency index which, however, does not measure inconsistency as well as uncertainty ideally. Here, we investigate the same problem once again and propose a new solution in more general setting.

2 Abelian linearly ordered groups

We shall investigate pairwise comparison matrices with elements being fuzzy quantities of the alo-group over an interval of the real line $\mathbf{R}$. Such an approach allows for unifying the theory dealing with additive, multiplicative and PCF matrices. The matter of this section is based mainly on [2], or [1].

An *abelian group* is a set, G, together with an operation $\odot$ (read: operation *odot*) that combines any two elements $a, b \in G$ to form another element denoted by $a \odot b$. The symbol $\odot$ is a general placeholder for a concretely given operation. The set and operation, $(G, \odot)$, satisfies the following requirements known as the *abelian group axioms*:

[1] Silesian University in Opava/School of Business Administration in Karvina, University Sq. 1934/3, 73340 Karvina, ramik@opf.slu.cz

- If $a, b \in G$, then $a \odot b \in G$ (*closure*).
- If $a, b, c \in G$, then $(a \odot b) \odot c = a \odot (b \odot c)$ (*associativity*).
- There exists an element $e \in G$ called the *identity element*, such that for all $a \in G$, $e \odot a = a \odot e = a$ (*identity element*).
- If $a \in G$, then there exists an element $a^{(-1)} \in G$ called the *inverse element to* a such that $a \odot a^{(-1)} = a^{(-1)} \odot a = e$ (*inverse element*).
- If $a, b \in G$, then $a \odot b = b \odot a$ (*commutativity*).

The *inverse operation* $\div$ to $\odot$ is defined for all $a, b \in G$ as follows: $a \div b = a \odot b^{(-1)}$.

A nonempty set G is *linearly (totally) ordered* under the order relation $\leq$, if the following statements hold for all $a, b, c \in G$:

- If $a \leq b$ and $b \leq a$, then $a = b$ (*antisymmetry*).
- If $a \leq b$ and $b \leq c$, then $a \leq c$ (*transitivity*).
- $a \leq b$ or $b \leq a$ (*totality*).

The *strict order* relation $<$ is defined for $a, b \in G$: $a < b$ if $a \leq b$ and $a \neq b$.

Let $(G, \odot)$ be an abelian group, G be linearly ordered under $\leq$. $(G, \odot, \leq)$ is said to be an *abelian linearly ordered group*, *alo-group* for short, if for all $c \in G$, if $a \leq b$ implies $a \odot c \leq b \odot c$.

If $\mathcal{G} = (G, \odot, \leq)$ is an alo-group, then G is naturally equipped with the order topology induced by $\leq$ and $G \times G$ is equipped with the related product topology. We say that $\mathcal{G}$ is a *continuous alo-group* if $\odot$ is continuous on $G \times G$.

Because of the associative property, the operation $\odot$ can be extended by induction to n-ary operation, $n > 2$. Then, for a positive integer n, the (n)-*power* $a^{(n)}$ *of* $a \in G$ is defined. We can extend the meaning of power $a^{(s)}$ to the case that s is a negative integer.

$\mathcal{G} = (G, \odot, \leq)$ is *divisible* if for each positive integer n and each $a \in G$ there exists the (n)-*th root of* a denoted by $a^{(1/n)}$, i.e. $\left(a^{(1/n)}\right)^{(n)} = a$. Moreover, the function $\|.\| : G \to G$ defined for each $a \in G$ by $\|a\| = \max\{a, a^{(-1)}\}$ is called a $\mathcal{G}$-*norm*. The operation $d : G \times G \to G$ defined by $d(a, b) = \|a \div b\|$ for all $a, b \in G$ is called a $\mathcal{G}$-*distance*.

Example 1. *Additive alo-group* $\mathcal{R} = (]-\infty, +\infty[, +, \leq)$ is a continuous alo-group with:
$e = 0,\ a^{(-1)} = -a,\ a^{(n)} = n.a$.

Example 2. *Multiplicative alo-group* $\mathcal{R}^+ = (]0, +\infty[, \bullet, \leq)$ is a continuous alo-group with:
$e = 1,\ a^{(-1)} = a^{-1} = 1/a,\ a^{(n)} = a^n$.

Example 3. *Fuzzy additive alo-group* $\mathcal{R}_a$=$(]-\infty, +\infty[, +_f, \leq)$, see [2], is a continuous alo-group with:

$$a +_f b = a + b - 0.5,\ e = 0.5,\ a^{(-1)} = 1 - a, a^{(n)} = n.a - \frac{n-1}{2}.$$

Example 4. *Fuzzy multiplicative alo-group* $]\mathbf{0}, \mathbf{1}[_{\mathbf{m}}$=$(]0, 1[, \bullet_f, \leq)$, see [2], is a continuous alo-group with:

$$a \bullet_f b = \frac{ab}{ab + (1-a)(1-b)}, e = 0.5, a^{(-1)} = 1 - a.$$

3 PC matrices with elements being fuzzy sets of alo-group over a real interval

Let G be a proper interval of the real line $\mathbf{R}$ and $\leq$ is the total order on G inherited from the usual order on $\mathbf{R}$, $\mathcal{G} = (G, \odot, \leq)$ be a real alo-group. We also assume that $\mathcal{G}$ is a divisible and continuous alo-group. Then G is an open interval, see [2].

Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be an $n \times n$ matrix where each element $\tilde{a}_{ij}$ of $\tilde{A}$ is a triangular (L, R)-fuzzy number with the membership function $\mu_{\tilde{a}_{ij}}, i, j \in \{1, 2, ..., n\}, i \neq j$, given as follows:

$$\mu_{\tilde{a}_{ij}}(x) = \begin{cases} L_{ij}^{-1}(x) & \text{if } x \in [L_{ij}(0), L_{ij}(1)], \\ R_{ij}^{-1}(x) & \text{if } x \in [R_{ij}(1), R_{ij}(0)], \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In most of the literature on fuzzy sets a triangular fuzzy number is defined by means of linear (L, R) functions, here we consider a more general case: strictly monotone functions or constant functions. Let L_{ij} be either a strictly increasing continuous function, $L_{ij} : [0, 1] \to \mathbf{R}$, or, L_{ij} be a constant function mapping interval $[0, 1]$ into a real number $y \in \mathbf{R}$. Similarly, R_{ij} is either a strictly decreasing continuous function, $R_{ij} : [0, 1] \to \mathbf{R}$, or R_{ij} is a constant function mapping interval $[0, 1]$ into a real number $y \in \mathbf{R}$. We assume that $L_{ij}(1) = R_{ij}(1)$. By L_{ij}^{-1}, or R_{ij}^{-1}, we denote the inverse functions of the increasing function L_{ij}, or, decreasing function R_{ij}, respectively. If L_{ij}, or R_{ij} is a constant function, then for each $y \in \mathbf{R}$ we define $L_{ij}^{-1}(y) = R_{ij}^{-1}(y) = 1$. The functions L_{ij} and R_{ij} are called the *left and right membership generating functions of* $\tilde{a}_{ij}$, respectively.

Moreover, for all $i \in \{1, ..., n\}$ we assume that $\mu_{\tilde{a}_{ii}}(x) = 1$ if $x = e$, and $\mu_{\tilde{a}_{ii}}(x) = 0$ otherwise. Here e is the *identity element of* $\mathcal{G}$. The matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ with the above mentioned triangular (L, R)-fuzzy number elements is called the *PCFN matrix*.

It is clear that each entry $\tilde{a}_{ij}$ of the PCFN matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ can be identified with a tripple $(a_{ij}^L; a_{ij}^M; a_{ij}^R)_{L_{ij}, R_{ij}}$, where L_{ij} and R_{ij} are the membership generating functions with the properties (1) and $a_{ij}^L = L_{ij}(0), a_{ij}^M = L_{ij}(1) = R_{ij}(1), a_{ij}^R = R_{ij}(0)$. For the sake of simplicity we shall omit the subscripts referring to functions L_{ij}, R_{ij}, i.e. we simply write $\tilde{a}_{ij} = (a_{ij}^L; a_{ij}^M; a_{ij}^R)$. Notice that the crisp numbers (non-fuzzy numbers) are special cases of triangular (L, R)-fuzzy numbers.

By the usual definition, the α-cuts $[\tilde{a}_{ij}]_\alpha = \{x \in \mathbf{R} | \mu_{\tilde{a}_{ij}}(x) \geq \alpha\}$ of $\tilde{a}_{ij}$ are closed convex subsets of G for all $\alpha \in]0, 1]$. The zero-cuts, $[\tilde{a}_{ij}]_0$, are defined as: $[\tilde{a}_{ij}]_0 = [L_{ij}(0), R_{ij}(0)]$. Then we define for all $i, j \in \{1, 2, ..., n\}, i \neq j$, and $\alpha \in [0, 1]$:

$$a_{ij}^L(\alpha) = \min\{a \in G \subset \mathbf{R} \mid a \in [\tilde{a}_{ij}]_\alpha\}, a_{ij}^R(\alpha) = \max\{a \in G \subset \mathbf{R} \mid a \in [\tilde{a}_{ij}]_\alpha\}. \tag{2}$$

Hence, the α-cut of $\tilde{a}_{ij}$ is a closed interval $[\tilde{a}_{ij}]_\alpha = [a_{ij}^L(\alpha), a_{ij}^R(\alpha)]$. Now, we shall define the concept of reciprocity and consistency properties for PCFN matrices.

Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be an $n \times n$ PCFN matrix, $\alpha \in [0, 1]$. $\tilde{A}$ is said to be α-$\odot$-*reciprocal*, if the following condition (C1) holds:
(C1) For every $i, j \in \{1, 2, ..., n\}$ and every $a_{ij} \in [\tilde{a}_{ij}]_\alpha$ there exists $a_{ji} \in [\tilde{a}_{ji}]_\alpha$ such that

$$a_{ij} \odot a_{ji} = e. \tag{3}$$

$\tilde{A}$ is said to be $\odot$-*reciprocal*, if condition (C1) holds for all $\alpha \in [0, 1]$.

Remark 1. If $\tilde{A} = \{\tilde{a}_{ij}\}$ is a PCFN matrix with crisp elements then condition (C1) coincides with the classical definitions of reciprocity.

Proposition 1. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a PCFN matrix, $\alpha \in [0, 1]$. $\tilde{A}$ is α-$\odot$-reciprocal, if and only if*

$$a_{ij}^L(\alpha) \odot a_{ji}^R(\alpha) = e \text{ and } a_{ij}^R(\alpha) \odot a_{ji}^L(\alpha) = e. \tag{4}$$

Using an $\odot$-reciprocal matrix, a consistency property of the PCFN matrix is necessary. Let the PCFN matrix $A = \{a_{ij}\}$ be crisp. Then we have the following definition, see e.g. [2]: $A = \{a_{ij}\}$ is $\odot$-*consistent* if for all $i, j, k \in \{1, 2, ..., n\}$

$$a_{ij} = a_{ik} \odot a_{kj}. \tag{5}$$

Then the following result holds, see e.g. [2].

Proposition 2. *$A = \{a_{ij}\}$ is $\odot$-consistent if and only if there exists a vector $w = (w_1, w_2, ..., w_n)$, $w_i \in G$ such that*

$$w_i \div w_j = a_{ij} \text{ for all } i, j \in \{1, 2, ..., n\}. \tag{6}$$

Now, we define consistency for PCFN matrices. Let $\alpha \in [0,1]$, $\tilde{A} = \{\tilde{a}_{ij}\}$ is said to be α-$\odot$-*consistent*, if the following condition (C2) holds:

(C2) For every $i,j,k \in \{1,2,...,n\}$, there exist $a_{ij} \in [\tilde{a}_{ij}]_\alpha$, $a_{ik} \in [\tilde{a}_{ik}]_\alpha$ and $a_{kj} \in [\tilde{a}_{kj}]_\alpha$ such that

$$a_{ij} = a_{ik} \odot a_{kj}.$$

$\tilde{A}$ is said to be $\odot$-*consistent*, if condition (C2) holds for all $\alpha \in [0,1]$.

Remark 2. Let $\alpha, \beta \in [0,1], \alpha \geq \beta$. If $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-consistent, then $\tilde{A}$ is β-$\odot$-consistent.

Let $\alpha \in [0,1]$, a vector $w = (w_1, w_2, ..., w_n)$, $w_i \in G$ for all $i \in \{1,2,...,n\}$, is an α-$\odot$-*consistent vector with respect to* $\tilde{A} = \{\tilde{a}_{ij}\}$ if for all $i,j \in \{1,2,...,n\}$ there exist $a_{ij} \in [\tilde{a}_{ij}]_\alpha$ such that

$$w_i \div w_j = a_{ij}. \tag{7}$$

Example 5. Consider $\odot = \cdot$, let $\tilde{C} = \{\tilde{c}_{ij}\}$ be a PCFN matrix with the membership generating functions for elements $\tilde{c}_{ij}$ given by L_{ij}, R_{ij}, where $1 \leq i < j \leq 3$. Here, for the symmetric elements, we assume that $L_{ji}(t) = (R_{ij}(t))^{-1}$ and $R_{ji}(t) = (L_{ij}(t))^{-1}$, where $t \in [0,1]$ and $1 \leq i < j \leq 3$. Matrix $\tilde{C}$ is given as follows:

$$\tilde{C} = \begin{pmatrix} 1 & (1;2;2) & (2;6;8) \\ (\frac{1}{2};\frac{1}{2};1) & 1 & (2;3;4) \\ (\frac{1}{8};\frac{1}{6};\frac{1}{2}) & (\frac{1}{4};\frac{1}{3};\frac{1}{2}) & 1 \end{pmatrix},$$

By Proposition 1, $\tilde{C}$ is $\cdot$ -reciprocal, i.e. α- $\cdot$ -reciprocal for all $\alpha \in [0,1]$. Moreover, $\tilde{C}$ is $\cdot$ -consistent, i.e. α- $\cdot$ -consistent for all $\alpha \in [0,1]$.

Moreover, let $\tilde{D} = \{\tilde{d}_{ij}\}$ be a PCFN matrix with the membership generating functions for elements $\tilde{d}_{ij}$: L_{ij}, R_{ij}, where $1 \leq i < j \leq 3$. Again, for the symmetric elements, we assume that $L_{ji}(t) = (R_{ij}(t))^{-1}$ and $R_{ji}(t) = (L_{ij}(t))^{-1}$, where $t \in [0,1]$ and $1 \leq i < j \leq 3$. Matrix $\tilde{D}$ is given as follows:

$$\tilde{D} = \begin{pmatrix} 1 & (1;2;2) & (7;8;9) \\ (\frac{1}{2};\frac{1}{2};1) & 1 & (2;3;3) \\ (\frac{1}{9};\frac{1}{8};\frac{1}{7}) & (\frac{1}{3};\frac{1}{3};\frac{1}{2}) & 1 \end{pmatrix}.$$

Here, $\tilde{D}$ is a 3×3 PCFN matrix with triangular fuzzy numbers as elements. Again, $\tilde{D}$ is $\cdot$ -reciprocal, however, $\tilde{D}$ is not $\alpha - \cdot$ -consistent as evidently (C2) is not satisfied for any $\alpha \in [0,1]$.

Remark 3. If $\tilde{A}$ is a PC matrix with crisp elements, i.e. $\tilde{A} = A$ and $\tilde{A}$ is $\odot$-consistent, then A is $\odot$-consistent in the sense of definition (5).

Proposition 3. *Let $\alpha \in [0,1]$. $\tilde{A}$ is α-$\odot$-consistent if and only if there exists a vector $w^\alpha = (w_1^\alpha, w_2^\alpha, ..., w_n^\alpha)$ with $w_i^\alpha \in G$ for all $i \in \{1,2,...,n\}$ such that*

$$w_i^\alpha \div w_j^\alpha \in [\tilde{a}_{ij}]_\alpha \text{ for all } i,j \in \{1,2,...,n\}, \tag{8}$$

or, equivalently

$$a_{ij}^L(\alpha) \leq w_i^\alpha \div w_j^\alpha \leq a_{ij}^R(\alpha) \text{ for all } i,j \in \{1,2,...,n\}. \tag{9}$$

Now, we mention some isomorphisms between two spaces of PCFN matrices on alo-groups of the real line. Particularly, reciprocity and consistency of PCFN matrices are saved by special transformations - fuzzy extensions of isomorphisms between two alo-groups. This result can be applied to all isomorphisms between the important alo-groups on the real line, see [2].

Proposition 4. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a PCFN matrix on G. Let $h : G \longrightarrow G'$ be a strictly increasing continuous isomorphism between two alo-groups $\mathcal{G} = (G, \odot, \leq)$ and $\mathcal{G}' = (G', \circ, \leq)$, $\alpha \in [0,1]$.*
(i) If $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-reciprocal, then $\tilde{A}' = \{\tilde{a}'_{ij}\} = \{\tilde{h}(\tilde{a}_{ij})\}$ is α-$\circ$-reciprocal, where $\tilde{h}$ is a fuzzy extension of h defined by

$$\mu_{\tilde{a}'_{ij}}(t) = \mu_{\tilde{a}_{ij}}(h^{-1}(t)), t \in G'. \tag{10}$$

(ii) If $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-consistent, then $\tilde{A}' = \{\tilde{a}'_{ij}\} = \{\tilde{h}(\tilde{a}_{ij})\}$ is α-$\circ$-consistent, where $\tilde{h}$ is a fuzzy extension of h defined by (10).

Finally, when at least for one triple of elements and some $\alpha \in [0,1]$, the condition (C2) is not satisfied, then the matrix $\tilde{A}$ is $\odot$-*inconsistent*. It is important to measure an intensity of $\odot$-inconsistency as in some cases the PCFN matrix can be "close" to an $\odot$-consistent matrix, or, in the other cases $\odot$-inconsistency can be strong, meaning that the PCFN matrix can be "far" from some $\odot$-consistent matrix.

4 Methods for deriving priorities from PCFN matrices

In this section, we assume that all matrices are $n \times n$ $\odot$-reciprocal PCFN matrices on $\mathcal{G} = (G, \odot, \leq)$, where G is an open interval in $]-\infty, +\infty[$. Here, definition of the priority vector for ranking the alternatives will be based on Proposition 3, particularly, the optimal solution of the following optimization problem will be useful.
(P1)

$$\alpha \longrightarrow \max; \tag{11}$$

subject to

$$a_{ij}^L(\alpha) \leq w_i \div w_j \leq a_{ij}^R(\alpha) \text{ for all } i, j \in \{1, 2, ..., n\}, i < j, \tag{12}$$

$$\bigodot_{k=1}^{n} w_k = e, \tag{13}$$

$$0 \leq \alpha \leq 1, w_k \in G, \text{ for all } k \in \{1, 2, ..., n\}. \tag{14}$$

If optimization problem (P1) has a feasible solution, i.e. system of constraints (12) - (14) has a solution, then (P1) has also an optimal solution. Let α^* and $w^* = (w_1^*, ..., w_n^*)$ be an optimal solution of problem (P1). Then $\alpha^* \geq 0$ and α^* is called the *$\odot$-consistency grade of $\tilde{A}$*, denoted by $G_\odot(\tilde{A})$, i.e.

$$G_\odot(\tilde{A}) = \alpha^*. \tag{15}$$

Here, $w^* = (w_1^*, ..., w_n^*)$ is an α^*-$\odot$-consistent vector with respect to $\tilde{A}$ called the *$\odot$-priority vector of* $\tilde{A}$.

If optimization problem (P1) has no feasible solution, then we define

$$G_\odot(\tilde{A}) = 0. \tag{16}$$

Problem (P1) is a nonlinear optimization problem that can be efficiently solved by the well known *dichotomy method*, which is a sequence of optimization problems, see e.g. [4]. For instance, given $\alpha \in [0, 1], \odot = +$, problem (P1) can be solved as an LP problem (with variables $w_1, ..., w_n$).

If optimization problem (P1) has no feasible solution, then $\tilde{A}$ is inconsistent and the $\odot$-consistency grade of $\tilde{A}$ is zero, i.e. $G_\odot(\tilde{A}) = 0$. The inconsistency of $\tilde{A}$ will be measured by the minimum of the $\odot$-mean distance of the *ratio matrix* $W = \{w_i \div w_j\}$ to matrix $\{[\tilde{a}_{ij}]_0\}$ where the elements are zero-cuts of $\tilde{a}_{ij}$, i.e. intervals $[\tilde{a}_{ij}]_0$.

Let $w = (w_1, ..., w_n), w_i \in G$ for all $i \in \{1, ..., n\}$. Denote

$$I_\odot(\tilde{A}, w) = \min\Big\{ \Big(\bigodot_{i<j} \|a_{ij} \div (w_i \div w_j)\|\Big)^{\left(\frac{2}{n(n-1)}\right)} \;\Big|\; a_{ij} \in [\tilde{a}_{ij}]_0 \text{ for all } 1 \leq i < j \leq n \Big\}.$$

Now, we define a priority vector in case there is no feasible solution of (P1). Clearly, such priority vector cannot become an α-$\odot$-consistency vector of $\tilde{A}$ for some $\alpha > 0$.

Consider the following optimization problem (P2).
(P2)

$$I_\odot(\tilde{A}, w) \longrightarrow \min; \tag{17}$$

subject to

$$\bigodot_{k=1}^{n} w_k = e, \tag{18}$$

$$w_k \in G, \text{ for all } k \in \{1, 2, ..., n\}. \tag{19}$$

The *$\odot$-consistency index of* $\tilde{A}$, $I_\odot(\tilde{A})$, is defined as

$$I_\odot(\tilde{A}) = I_\odot(\tilde{A}, w^*) \tag{20}$$

where $w^* = (w_1^*, ..., w_n^*)$ is the optimal solution of (P2).

If $G_\odot(\tilde{A}) = 0$, then the optimal solution of (P2) is called the *$\odot$-priority vector of* $\tilde{A}$.

Remark 4. If $\odot$-consistency grade $G_{\odot}(A) > 0$, then the $\odot$-priority vector of $\tilde{A}$ is unique. However, if $G_{\odot}(A) = 0$, then the uniqueness of optimal solution of (P2) is not saved which is an unfavorable fact from the point of view of the DM.

Example 6. Let $\tilde{E}$ be a PCFN matrix on the additive alo-group $\mathcal{R}$, i.e. $\odot = +$ as follows:

$$\tilde{E} = \begin{pmatrix} 0 & (1;2;2) & (4;8;9) \\ (-2;-2;-1) & 0 & (3;4;4) \\ (-9;-8;-4) & (-4;-4;-3) & 0 \end{pmatrix}.$$

Here, $\tilde{E}$ is a $+$-reciprocal PCFN matrix, the elements of which are triangular (L,R)-fuzzy numbers and all membership generation functions $L_{ij}, R_{ij}, i \neq j$, are supposed to be linear. Then the consistency grade $G_+(\tilde{E}) = 0.4$, and the $+$-priority vector of $\tilde{E}$ is $w^* = (2.533, 0.533, -3.067)$. The consistency index $I_+(\tilde{E}) = 1.333$, hence, $\tilde{E}$ is $+$-inconsistent.

5 Conclusion

This paper investigated pairwise comparison matrices with fuzzy elements. In comparison to pairwise comparison matrices with crisp elements investigated in the literature, here we investigated pairwise comparison matrices with elements from abelian linearly ordered group (alo-group) over a real interval. We generalized the concept of reciprocity and consistency of pairwise comparison matrices with triangular fuzzy numbers (PCFN matrices). Moreover, we introduced the concept of priority vector which is a generalization of the crisp concept. Such an approach allows for a generalization dealing both with the PCFN matrices on the additive, multiplicative and also fuzzy alo-groups. We also unify several approaches known from the literature, see e.g. [3], [4], [6], [9], and [5]. The problem of measuring the inconsistency of PCFN matrices is solved by defining corresponding indexes. Numerical examples were presented to illustrate the concepts and derived properties.

Acknowledgements

This research has been supported by GACR project No. 14-02424S

References

[1] Bourbaki, N.: *Algebra II.* Springer Verlag, Heidelberg-New York-Berlin, 1990.

[2] Cavallo, B., D'Apuzzo, L., Squillante, M.: About a Consistency Index for Pairwise Comparison Matrices over a Divisible Alo-Group. *International J. of Intelligent Systems* **27** (2012), 153–175.

[3] Mikhailov, L.: Deriving priorities from fuzzy pairwise comparison judgments. *Fuzzy Sets and Systems* **134** (2003), 365–385.

[4] Ohnishi, S., Dubois, D. at al.: A fuzzy constraint based approach to the AHP. In: *Uncertainty and Intelligent Inf. Syst.*, World Sci., Singapore, (2008), 217–228.

[5] Ramik, J., Korviny, P.: Inconsistency of pairwise comparison matrix with fuzzy elements based on geometric mean. *Fuzzy Sets and Systems* **161** (2010), 1604–1613.

[6] Salo, A.A.: On fuzzy ratio comparison in hierarchical decision models. *Fuzzy Sets and Systems* **84** (1996), 21-32.

[7] Van Laarhoven, P.J.M. and Pedrycz, W.: A fuzzy extension of Saaty's priority theory. *Fuzzy Sets and Systems* **11** (1983), 4, 229–241.

[8] Vaidya, O. S. and Kumar, S.: Analytic hierarchy process: An overview of applications. *European Journal of Operational Research* **16** (2006), 9, 1, 1–29.

[9] Xu, Z.S., Chen, J.: Some models for deriving the priority weights from interval fuzzy preference relations. *European Journal of Operational Research* **184** (2008), 266–280.

RBC model vs. New Keynesian model: time series analysis of Czech data

Vlastimil Reichel[1], Miroslav Hloušek[2]

Abstract. The aim of this paper is to analyse the macroeconomic data of Czech economy and to discuss whether they correspond to RBC or New Keynesian theory. Firstly, we focus on non-neutrality of monetary policy. Secondly, using Blanchar-Quah decomposition in bivariate VAR model we identify technology and nontechnology shocks. The technology shocks produces negative correlation between hours worked and productivity, while nontechnology shocks implies positive correlation between these two variables. This result is in contrast with RBC model which shows positive correlation between hours worked and productivity in reaction to technology shocks, as main source of fluctuations. On the other hand, New Keynesians models are consistent with these empirical findings.

Keywords: VAR model, Real Business Cycle Theory, New Keynesian Theory

JEL classification: E32

1 Introduction

Business cycles are still central topic of macroeconomics and macroeconomic modeling. There are many theories that try to explain business cycle fluctuations. Goal of this paper is to decide which of the two approaches – Real Business Cycle (RBC) theory or New Keynesian (NK) theory – is more in accordance with Czech economy data. This paper closely follows works of Galí [3] and Galí [4]. The analysis is made on quarterly data of Czech economy covering period 1996:Q1-2012:Q4. We use structural VAR models for processing of data and basic DSGE models as representatives of the theories. First, we focus on identification of non-neutrality of monetary policy. This assumption is tested by impulse responses and historical shock decomposition obtained from the estimated structural VAR model. Second, we compare data characteristics obtained from bivariate VAR model with results from DSGE models. The tools are again impulse response functions and conditional and unconditional correlations calculated from simulated series.

The results show that there is some evidence of monetary policy non-neutrality in Czech data which favours New Keynesian theory. This theory is also supported by empirical fact that technology shock produces negative correlation between productivity and worked hours while nontechnology shock shows negative correlation between these variables. Real business cycle theory fails to replicate these findings.

2 Evidence of Monetary Policy Non-neutralities

According to RBC theory the monetary policy is neutral in the short run. On the other hand, NK theory assumes short run non-neutrality of monetary policy. This non-neutrality is a natural consequence of the presence of nominal rigidities. If prices do not adjust in proportion to changes in the money supply; or if expected inflation does not move along with the nominal interest rate when the latter is changed, the central bank would generally be able to change the level of demand and, as a result, the equilibrium levels of output and employment.

To identify structural monetary shock we estimated structural VAR model and imposed Choleski decomposition. In this model we work with six variables: GDP (y), worked hours (h), deflator of GDP (def), nominal interest rate (ir), monetary aggregate M2 ($M2$), and labor productivity (p).[1]. All variables are in logarithms, except of nominal interest rate. We use Hodrick Prescott filter to obtain the cyclical component. Ordering of the variables is as follows $x_t = [\hat{ir}, \hat{def}, \hat{M2}, \hat{y}, \hat{p}, \hat{h}]'$.

[1]Masaryk University, Lipová 41a, Brno, Czech Republic, e-mail: 357467@mail.muni.cz
[2]Masaryk University, Lipová 41a, Brno, Czech Republic, e-mail: hlousek@econ.muni.cz
[1]Labor productivity is calculated as the ratio of GDP and total employment

First, we look at impulse responses. Figure 1 shows the reaction of the variables to monetary policy shock – increase of the nominal interest rate by one standard deviation. The solid line represents point estimates of impulse response functions, the dashed lines capture corresponding 90 % confidence interval. On the x-axis we measure the quarters after the shock, the y-axis shows percentage deviation from trend.

Interest rate increases by 0.75 percentage points and then slowly returns to the original level. GDP deflator at first increases and then gradualy declines to steady state. This initial increase is quite counterintuitive and is often favours as price puzzle. It was found also in other studies and is generally attributed to the existence of nominal rigidities (see references in Walsh [7]). Reaction of M2 is largely negative and is connected with liquidity effect – central bank must bring down money supply to increase interest rate. Primal subject of our interest are the real variables. GDP increases and then declines with a characteristic hump-shaped pattern. Reaction of productivity and worked hours move hand in hand with the reaction of GDP. Initial increase of the real variables in reaction to contractionary policy shock is again puzzling. It can be ascribed to existence of real rigidities as habit in consumption or long-term contracts on investments.

We can see that monetary policy shock has some real effects in the short run. These findings are partly reduced if we take into account the uncertainty connected with the estimates. However, there are still several periods where confidence bands of impulse responses do not include zero. Thus it can be interpreted as evidence of non-neutrality of monetary policy.

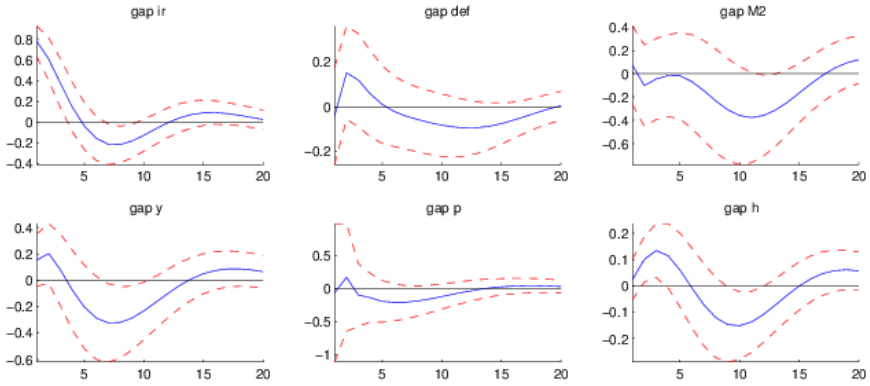


Figure 1: VAR model: Impulse responses to shock ir

Another way to examine the impact of monetary shocks to the real variables is the historical decomposition. Figure 2 shows decomposition of the GDP gap into individual shocks obtained from the estimated SVAR model. Nominal shocks are depicted by shades of blue color and the real shocks are painted in other colors.

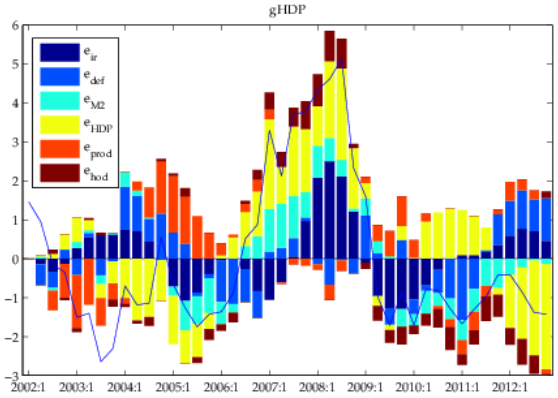


Figure 2: SVAR model - Historical shock decomposition

The effects of monetary shocks (represented by the shocks in $M2$ and ir) are quite significant. For example, these two shocks importantly contributed to the boom between 2006 and 2009. However, negative impact of these shocks is also visible during recent recession. Average contributions of individual shocks to behavior of the GDP

gap are quoted in Table 1. We can see that the average impact of monetary shocks accounts for 38 % of the total variance of the GDP gap.

	monetary		other			
shocks	e_{ir}	e_{M2}	e_{def}	e_{HDP}	e_{prod}	e_h
values (in %)	19.08	18.72	13.00	24.71	15.77	8.72

Table 1: The average percentage contribution of individual shocks

3 Data fit of the models

In this section we compare data characteristics obtained from bivariate VAR model with Blanchard-Quah decomposition and statistics from simulation of DSGE models. We focus on two variables – productivity and worked hours and two specifications of bivariate VAR models (theses specifications follow Galí [3]): $[\Delta p_t, \Delta h_t]'$ and $[\Delta p_t, \hat{h}_t]'$, where Δp_t is the first difference of productivity, Δh_t is the first difference of worked hours and $\hat{h}_t$ is the gap of worked hours. Blanchar-Quah decomposition helps to identify two types of structural shocks. Shocks that have permanent impact on productivity are called technology (or supply) shocks, shocks that have only temporary impact are non-technology (or demand) shocks (see Blanchard and Quah [1] or Enders [2]).

We will look at two measures: impulse responses and correlations. The representative of RBC theory is Hansen's model [5] and New Keynesian model comes form Ireland [6].

3.1 Models

We present the models in log-linearized form, the variables are expressed as deviation from steady-state. For detailed derivations see the original papers. Hansen's RBC model is comprised of six equations for following six variables: consumption $\hat{c}$, worked hours $\hat{h}$, interest rate $\hat{r}$, output $\hat{y}$, capital $\hat{k}$ and technology shock $\hat{z}$. The model equations are:

$$\hat{c}_t = E_t\hat{c}_{t+1} - \beta\bar{r}E_t\hat{r}_{t+1} \tag{1}$$

$$\hat{y}_t = \frac{\hat{h}_t}{1-\bar{h}} + \hat{c}_t \tag{2}$$

$$\hat{y}_t\bar{y} = \hat{c}_t\bar{c} - \bar{k}\left[(1-\delta)\hat{k}_t - \hat{k}_{t+1}\right] \tag{3}$$

$$\hat{z}_t = \hat{y}_t - \alpha\hat{k}_t - (1-\alpha)\hat{h}_t \tag{4}$$

$$\hat{r}_t = \hat{y}_t - \hat{k}_t \tag{5}$$

$$\hat{z}_t = \rho\hat{z}_{t-1} + \varepsilon_t \tag{6}$$

The equations (1) and (2) are Euler and intratemporal equation, respectively, and were obtained from utility maximization problem of households. Equation (3) is economy budget constraint, (4) is the production function and (5) is the marginal product of capital. The last equation is shock to technology. Calibrated parameters are listed in Table 2.

Parameters:	α	β	δ	ρ_z	σ_z
Values:	0.4	0.98	0.05	0.95	0.0035

Table 2: RBC model - calibrated parameter values

New Keynesian model of Ireland [6] is characterized by eight equations for the following eight variables: output $\hat{y}$, inflation $\hat{\pi}$, worked hours $\hat{h}$, money $\hat{m}$, nominal interest rate $\hat{r}$, aggregate technology shock $\hat{z}$, preference shock $\hat{a}$, and money demand shock $\hat{e}$. The model equations are:

$$\hat{y}_t = \hat{y}_{t+1} - (\hat{r}_t - \hat{\pi}_{t+1}) + \omega((\hat{m}_t - \hat{e}_t) - (\hat{m}_{t+1} - \hat{e}_{t+1})) + (\hat{a}_t - \hat{a}_{t+1}) \tag{7}$$

$$\hat{m} = \gamma_1 \hat{y} - \gamma_2 \hat{r} + \gamma_3 \hat{e} \tag{8}$$

$$\hat{\pi} = \beta \hat{\pi}_{t+1} + \psi(\hat{y}_t - \omega(\hat{m}_t - \hat{e}_t) - \hat{z}_t) \tag{9}$$

$$\hat{r}_t = \rho_r \hat{r}_{t-1} + \rho_y \hat{y}_{t-1} + \rho_\pi \hat{\pi}_{t-1} + \varepsilon_m \tag{10}$$

$$\hat{z}_t + \hat{h}_t = \hat{y}_t \tag{11}$$

$$\hat{z}_t = \rho_z \hat{z}_{t-1} + \varepsilon_{zt} \tag{12}$$

$$\hat{e}_t = \rho_e \hat{e}_{t-1} + \varepsilon_{et} \tag{13}$$

$$\hat{a}_t = \rho_a \hat{a}_{t-1} + \varepsilon_{at}. \tag{14}$$

Equation (7) is the IS curve, (8) represents the demand for money, equation (9) is the Phillips curve, equation (10) is the Taylor rule, equation (11) is the production function and equations (12) to (14) are technology, money demand and preference shocks, respectively. Model calibration is listed in Table 3.

Parameters	β	ω	ψ	γ_1	γ_2	γ_3	ρ_r	ρ_y
Values:	0.98	0.25	0.2	0.0158	0.1251	0.9977	0.8	0.5
Parameters	ρ_π	ρ_a	ρ_e	ρ_z	σ_a	σ_e	σ_z	σ_r
Values:	1.5	0.8	0.8	0.8	0.01	0.01	0.01	0.01

Table 3: NK model - calibrated parameter values

3.2 Results

Firstly, we focus on a comparison of impulse responses. Figures 3 and 4 shows response of productivity and hours worked to technology and non-technology shocks identified in data using the VAR model with Blanchard Quah decomposition.

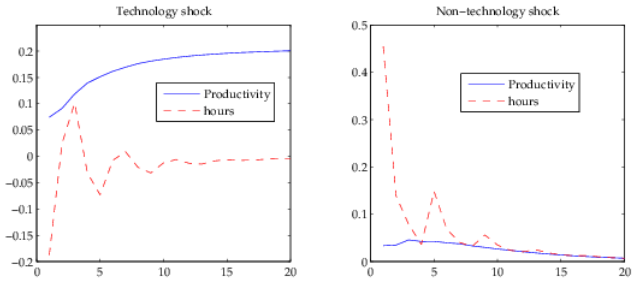


Figure 3: VAR model $[\Delta p_t, \Delta h_t]'$: Impulse responses to technology and non-technology shocks

In reaction to technology shock we see positive and permanent response of productivity and initial negative response of worked hours. Considering the non-technology shock, the initial reaction of productivity and worked hours is positive. This is evident in both figures, for both detrending methods.

Figure 5 shows reaction of output and worked hours to technology shock (which is the only source of fluctuations) for the RBC model.[2] We observe a positive response in both variables. This result is in stark contrast to reaction in the data.

Finally, Figure 6 shows reaction of the variables to technology and non-technology shock in New Keynesian model. We see similar initial reaction as in the data (Figures 3-4). Depending on the technology shock we find a negative initial response of worked hours and positive initial response of output; and depending on the non-technology shock we find positive initial response of both variables. Impulse responses obtained from Ireland's New Keynesian model corresponds with the responses in data.

[2]The variable *output* in the models denotes output per capita and it corresponds to variable *productivity* considered in data.

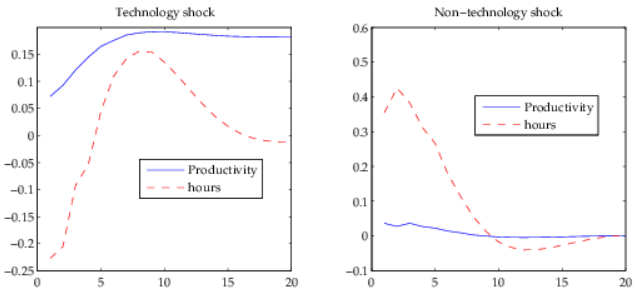


Figure 4: VAR model $[\Delta p_t, \hat{h}_t]'$: Impulse responses to technology and non-technology shocks

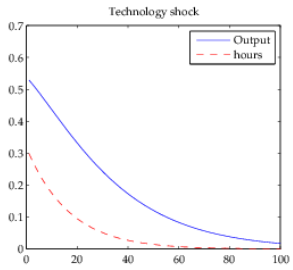


Figure 5: RBC model: Impulse responses to technology shocks

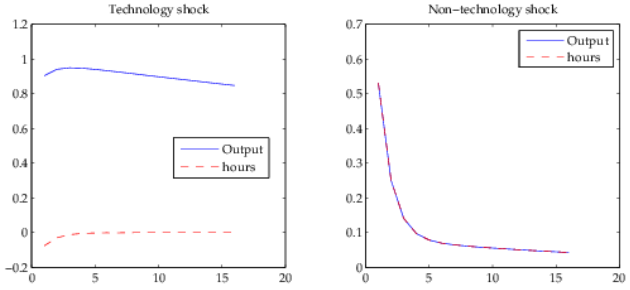


Figure 6: NK model: Impulse responses to technology and non-technology shocks

Now we focus on the correlations. Table 4 reports the unconditional and conditional correlations between the worked hours and productivity. The conditional correlations are obtained from simulations of VAR model and DSGE models with contribution of only technology or non-technology shocks. The first two panels reports correlations from VAR model for both specifications: Data A $[\Delta p_t, \Delta h_t]'$ and Data B $[\Delta p_t, \hat{h}_t]'$.

Unconditional correlation between worked hours and productivity (the first column) are small and negative or positive (depending on the detrending procedure). The conditional correlations of technology component are negative and large in both cases (the second column) and the conditional correlations of non-technology components are positive and small (the third column). When we compare the correlations from the RBC model with the data, we do not find a match. The correlation conditioned by technology shock is positive and large (0.67). Because there is only technology shock in the model the unconditional correlation is the same as the conditional correlation of technology shock.

Last panel shows correlations for New Keynesian model. The unconditional correlation is small and positive (0.22). This corresponds to unconditional correlation calculated from data (panel Data A). Correlation conditioned by technology shock is negative and large (-0.77) which is again in accordance with data. Correlation conditioned by non-technology shocks is positive and very high (0.99). The sign coincides with data but the size is much higher. The reason could be the simplicity of the model.

	Unconditional	**Conditional**	
		Techn. comp.	Non-techn. comp.
Data A: Δp			
Hours Δh	0.16	-0.64	0.33
Data B: Δp			
Hours $\hat{h}$	-0.12	-0.73	0.01
RBC model			
Hours	0.67	0.67	-
NK model			
Hours	0.22	-0.77	0.99

Table 4: Data, RBC, NK - Estimates of conditional and unconditional correlations

4 Conclusion

In this paper, we tested two theories of business cycles – RBC and New Keynesian theory – in terms how they fit the data of Czech economy. In the first part, we focused on assumptions of the theories, concretely existence of non-neutrality of monetary policy. We found out that monetary policy have some short run impacts on behaviour of the real variables. This conclusion comes from impulse responses to monetary policy shock and historical shock decomposition of GDP gap.

In the second part, we compared the characteristics of the data with the model simulations. The characteristics of the data were obtained from estimated bivariate VAR model with imposed Blanchard-Quah decomposition. Specifically, we identified technology and nontechnology shocks and studied their impacts for behaviour of productivity and worked hours. The results were as follows: estimated conditional correlations of hours and productivity are negative for the technology shock and positive for non-technology shock. The results from simulation of New Keynesian model are consistent with these results. On the contrary, RBC model shows positive conditional correlation between worked hours and productivity for technological shock. These findings were also supported by impulse responses. Thus, New Keynesian corresponds to the Czech economy data better than the RBC model.

Acknowledgements

This paper is supported by specific research project no. MUNI/A/0775/2013 at Masaryk University.

References

[1] Blanchard, Quah. *The Dynamic Effects of Aggregate Demand and Supply Disturbances*, American Economic Rewiew **79** (1989), pp. 655–673.

[2] Enders, Walter. *Applied Econometric Times Series*. USA: University of Alabama, 2010.

[3] Galí, Jordi. *Technology, Employment, and the Business Cycle: Do technology Shocks Explain Aggregate Fluctulations?* American Economic Rewiew **249** (1999).

[4] Galí, Jordi. *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework.* Princeton: Princeton University Press, 2008.

[5] Hansen, G. *Indivisible labor and the business cycle.* Journal of Monetary Economics **16** (1985), 309–327. North-Holland.

[6] Ireland, Peter N. *Technology Shocks in the New Keynesian Model*. Review of Economics and Statistics **86** (2004), 923–936.

[7] Walsh, C. E. *Monetary Theory and Policy.* Cambridge: The MIT Press, 2nd edition, 2003.

Factor and Principal Component Analysis in Scoring Modeling using SAS® System

Michal Řičař[1]

Abstract. Scoring models are used to evaluate probability of negative event. Banks and nowadays corporate segment in general also use scoring models to decrease risk of their business. The scoring model building procedure has been highly developed and currently it is possible to find best practices which are highly effective in the building process. The typical situation in the building process is that we dispose with large database of variables with thousands or millions of observations. Due to main econometrics requirements, such as multicollinearity has to be low, we are forced to use techniques which find out best explanatory variables between hundreds of variables without significant information loss. Factor and Principal Component Analysis are instruments which can be used to reveal strongest and most useful variables with respect to their information importance. In this article we used SAS® System to apply mentioned techniques on large data set. Our findings show that Factor and Principal Component Analysis are very effective instruments which are able to divide variables with respect to whole set of information, thus we are able significantly improve our modeling.

Keywords: Factor analysis, Principal Component Analysis, Cluster Analysis, Scoring, SAS.

JEL Classification: C49
AMS Classification: 91B82

1 Introduction

During scoring modeling we are challenged by many alternative ways how we can build a scoring model. Explanatory variables are one of the main issues that modeler faces – their connections, multicollinearity. Common scoring modeling situation offers tens or hundreds of variables which have economic sense but from econometric point of view they are problematic in the modeling procedure. The main problem is multicollinearity between variables which causes insignificant coefficients of variables in the model and coefficients interpretation has no information. Whereas scoring models are built for corporate purpose it is necessary to dispose with variables and their coefficients which have economic and logic interpretation. The solution could provide factor analysis (FA) and principal component analysis (PCA).

2 Ordination analysis

Multidimensional data set can be simplified by the FA or PCA by revealing main trends in variability of full set of variables. These methods are included in ordination (gradient) analysis. In general, the ordination analysis goes as follows - we order objects by ordination axes with respect to data structure and trend. Objects are points in multidimensional space where number of dimensions is equal to number of variables. An object position is determined by values of variables. Thus, the success of ordination analysis depends on data structure. Well-structured data (correlated variables) provides concentration of main part of variability to few ordination axes, which are usually orthogonal.

The purpose of ordination analysis is compression of huge data set, which is problematic for interpretation, to small range of variables (factors, components) without significant loss in information from original data set. Further, obtained factors (components) summarize groups of variables which have similar variability; hence these factors (components) can provide correlation structure of observed variables. From economic point of view, factors (components) represent latent characteristics of original variables and summarize them.

The ordination methods are useful exploration techniques which don't consider casual relationships, thus, they are not suitable for testing hypothesis but rather to create them.

All of the ordination methods, which reduce dimensions, are based on *eigenanalysis*, i.e. searching for *eigenvectors* and *eigenvalues* of association matrix. Association matrix could be covariance or correlation matrix. In the following paragraph we will shortly describe formal logic.

[1]University of Economics, W. Churchill Sq. 4, Prague, Czech Republic, e-mail: michal.ricar@vse.cz

2.1 Mathematical background

Let a matrix **A** be a square symmetrical matrix. We are looking for a matrix **Λ** which is equivalent to the matrix **A** and which has non-zero values only on its diagonal. Let the matrix **Λ** be called as a matrix of eigenvalues (also known as a canonical form of the matrix **A**).

Eigenvalues and eigenvectors of the matrix **A** can be found by

$$\boldsymbol{A}\boldsymbol{u}_j = \lambda_j \boldsymbol{u}_j, \tag{1}$$

thus we obtain eigenvalues λ_j and eigenvectors $\mathbf{u}_j$. Equation (1) can be rewritten as

$$(\boldsymbol{A} - \lambda_j \boldsymbol{I})\boldsymbol{u}_j = 0. \tag{2}$$

Besides the basic solution, where $\boldsymbol{u}_j$ is zero vector, the equation (2) has non-trivial solution

$$\left|\boldsymbol{A} - \lambda_j \boldsymbol{I}\right| = 0, \tag{3}$$

i.e. the determinant of the difference between matrices $\boldsymbol{A}$ a $\lambda_j \boldsymbol{I}$ has to be equal to zero for each λ_j, which is called as a characteristic equation.

The characteristic equation for the matrix **A** of order *p* is equal to a polynomial λ of order *p*. After we compute eigenvalues by (3) we are able to compute eigenvectors simply by substituting eigenvalues to (2).

Since the matrix **A** is a symmetrical matrix its eigenvectors are orthogonal.

2.2 Interpretation of results of ordination analysis

Mathematical meaning of ordination analysis can be interpreted as a set of orthogonal vectors which rearrange variability of original variables to express the maximum of this variability in minimum number of dimensions. Besides mathematical meaning we can find different views which are connected with the purpose of an analysis:

- Simplify multidimensional data set to minimum number of dimensions without interpretation needs, thus we only want to look at the data set in few dimensions which could describe most of the original variability but we don't need it for further computations as variables nor for correlation understanding (only for data reduction needs) – then PCA is a good choice,
- Simplify multidimensional data set to minimum number of dimensions with interpretation needs, hence we need to use a small range of variables, which are independent, and contains most of variability of original variables, and these dimensions (factors or components) have economic sense – then FA is a good choice,
- Identify correlation structures of original variables. Each factor (component) is composed by few or all original variables thus we are able to identify their connections and logic (economic) pattern – then FA is a good choice.

2.3 Common Factor Analysis and Principal Component Analysis

Common factor analysis was invented by Spearman (1904). Kim and Mueller (1978a, 1978b) provide a very elementary discussion of the common factor model. Gorsuch (1974) presents a broad survey of factor analysis, and Gorsuch (1974) and Cattell (1978) are useful as guides to practical research methodology [3] (literature can be found in [3]). Principal component analysis was originated by Pearson (1901) and later developed by Hotelling (1933). The application of principal components is discussed by Rao (1964), Cooley and Lohnes (1971), and Gnanadesikan (1977). Excellent statistical treatments of principal components are found in Kshirsagar (1972), Morrison (1976), and Mardia, Kent, and Bibby (1979) [4] (literature can be found in [4]).

FA and PCA offer latent variables which significantly simplify original set of variables. Orthogonal latent variables provide perfect characteristics for scoring modeling which is based on Bayes' theorem – orthogonal explanatory variables. On the other hand, business requirements don't allow using latent variables directly due to their non-direct economic interpretation. Hence, in many cases we are forced to use FA or PCA only for data structure view (the most important variables in data variability) and correlation structures between variables.

The PCA algorithm goes as follows – for the correlation or covariance matrix of original data set the method finds the first main factor which explains biggest part of data set's variability. Next factors are computed as orthogonal with a previous one with respect to exhaust maximum variability in descending order. After *p* steps PCA

finds all of the components and explains full variability in the data set. First few components are able to explain most of the variability, thus we are able to significantly reduce number of dimensions.

FA is based on the same logic but a different approach is used. It is assumed that the variance of a single variable can be decomposed into common variance which is shared by other variables and unique variance which is unique to a particular variable and includes the error component. FA, instead of PCA, analyzes only the common variance [2]; PCA considers total variance thus doesn't distinct between common and unique variance. FA goes further and in next step decides how many factors should be kept and interpreted (our decision about number of factors should be based on criteria like Kaiser-Guttman rule, Percentage of Variance, Scree Test, Analysis of Residuals and Interpretability). Finally, in the third step, after we decided how many factors should be kept, FA rotates factors in the way which helps us with their interpretation. Factors are rotated to catch most of similar variables; thus each variable should have big factor loadings with small range of conjoint factors and small or middle factor loadings with the rest of factors (factor loadings - correlation coefficients of variables with their factors).

The fundamental theorem of factor analysis is invariant within rotations. That is, the initial factor pattern matrix is not unique. We can get an infinite number of solutions, which produce the same correlation matrix, by rotating the reference axes of the factor solution to simplify the factor structure and to achieve a more meaningful and interpretable solution. The idea of simple structure has provided the most common basis for rotation, the goal being to rotate the factors simultaneously so as to have as many zero loadings on each factor as possible [2]. The Rotation of factors exists in two alternatives. Orthogonal or non-orthogonal rotation, whereas common orthogonal rotation is varimax (variance maximizing), which maximize sum of variance of all factors, and quartimax, which minimize number of factors which are needed to explain a whole set of original variables.

The selection of one technique over the other is based upon several criteria. In common factor analysis, a small number of factors are extracted to account for the correlations among the observed variables - to identify the latent dimensions that explain why the variables are correlated with each other. Hence, FA is designed to explain why certain variables are correlated. In principal component analysis, the objective is to account for the maximum portion of the variance present in the original set of variables with a minimum number of composite variables called principal components [2].

Important criteria are also assumptions about the variance in the original variables. If the observed variables are measured relatively error free, (for example, age, years of education, or number of family members), or if it is assumed that the error and specific variance represent a small portion of the total variance in the original set of the variables, then principal component analysis is appropriate. But if the observed variables are only indicators of the latent constructs to be measured (such as test scores or responses to attitude scales), or if the error (unique) variance represents a significant portion of the total variance, then the appropriate technique to select is common factor analysis [2].

3 Modeling

We mentioned earlier, PCA doesn't distinct between common and unique parts of the variance presents in a variable. Hence, original correlation (covariance) matrix is submitted to an analysis with value 1 on the main diagonal. The second technique, FA, uses variance which is both error free and shared with other variables, this is called as prior communality estimates. Thus, in the first step FA substitutes the diagonal of the correlation matrix with communality estimates. Since communality estimates is strictly theoretical, we have to decide what technique should be used to obtain it. SAS offers many options which could help us – from very simple one which uses the largest absolute correlation for a variable with any other variable as the communality estimate for the variable; to more sophisticated approach that uses the squared multiple correlation between the variable and all other variables etc. (for more variety approaches see [1]).

Since both methods are different, we demonstrate their results to assess their usability on same data set of financial variables. The full data set has been provided by Bisnode for more than 100 thousand firms and we used 17 financial indicators from 4 groups (profitability, liquidity, activity and stability) as variables.

3.1 Common Factor Analysis

SAS offers many options with PROC FACTOR procedure. The most important options are METHOD - specifies the method for extracting factors; PRIORS - specifies a method for computing prior communality estimates; and ROTATE - specifies the rotation method. We used two PRIORS options and two ROTATE options to assess results from different approaches.

Configurations are:

- METHOD=P PRIORS=SMC,
- METHOD=P PRIORS=MAX,
- METHOD=P PRIORS=SMC ROTATE=VARIMAX,
- METHOD=P PRIORS=SMC ROTATE= PROMAX,

where
method P uses same procedure as Principal Component Analysis (extracting factors to be the principal-axis factoring method),
priors SMC sets the prior communality estimate for each variable to its squared multiple correlation with all other variables,
priors MAX sets the prior communality estimate for each variable to its maximum absolute correlation with any other variable,
rotate VARIMAX maximize sum of variance of all factors,
rotate PROMAX specifies oblique (i.e. non-orthogonal) promax rotation.

Figure 1 clearly shows that 5 factors are able to explain almost all of the variability of 17 observed variables. This result, in fact, is not affected by configuration of the factor procedure, thus it is robust solution for further interpretation.

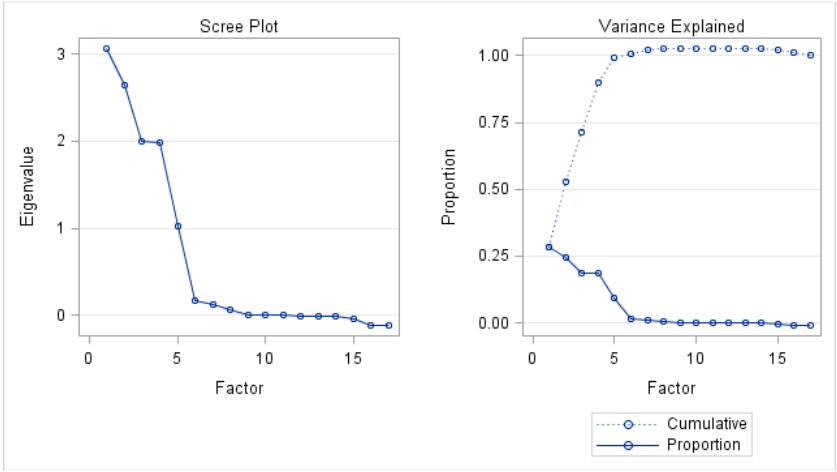


Figure 1 Scree Plot and Variance Explained of FA

Table 1 summarizes factor patterns for two types of prior communality estimates – SMC and MAX. The left table shows original values whereas the right table relative differences from the left table under the MAX option. We can clearly see that the both approaches are almost similar in the case of this data set.

METHOD=P PRIORS=SMC

Factor Pattern					
	Factor1	**Factor2**	**Factor3**	**Factor4**	**Factor5**
FI_1001	0.**74055**	-0.00293	-0.06804	-0.00452	-0.**54301**
FI_1002	0.**58541**	-0.00259	-0.01062	-0.00074	0.**54400**
FI_1003	0.08263	-0.00182	0.**99288**	0.08294	-0.02181
FI_1004	0.00348	0.**76324**	0.00082	0.00094	0.00035
FI_1005	0.00348	0.**98949**	0.00077	0.00167	-0.00014
FI_1006	0.00356	0.**98914**	0.00131	0.00171	-0.00014
FI_1008	0.00001	0.00025	-0.00003	0.00036	-0.00000
FI_1009	-0.00188	-0.00211	-0.08378	0.**99250**	-0.00013
FI_1011	-0.00100	-0.00200	-0.08295	0.**99257**	0.00094
FI_1013	0.00214	0.31756	0.00007	0.00087	0.00003
FI_1015	-0.00059	0.00149	-0.00024	-0.00026	-0.00001
FI_1016	0.00220	0.00335	-0.00875	-0.00046	-0.00517
FI_1017	-0.00110	0.00009	-0.00064	0.00047	-0.00087

METHOD=P PRIORS=MAX

Factor Pattern					
	Factor1	**Factor2**	**Factor3**	**Factor4**	**Factor5**
FI_1001	0,7%	-19,7%	-0,3%	-46,2%	5,1%
FI_1002	-3,4%	-21,3%	-14,6%	-52,6%	-8,2%
FI_1003	-2,7%	0,0%	0,7%	-40,2%	-1,4%
FI_1004	-21,6%	-5,3%	7,9%	-3,1%	-2,8%
FI_1005	-21,3%	0,9%	28,3%	6,4%	0,0%
FI_1006	-20,9%	0,9%	15,9%	4,3%	7,7%
FI_1008	0,0%	4,2%	-40,0%	2,9%	-
FI_1009	-3,1%	5,5%	-40,1%	0,5%	0,0%
FI_1011	-5,7%	5,8%	-40,3%	0,5%	1,1%
FI_1013	-18,9%	-8,4%	-	-8,4%	50,0%
FI_1015	0,0%	-1,3%	4,3%	-3,7%	-
FI_1016	-0,9%	-0,3%	1,5%	-51,1%	-28,6%
FI_1017	0,0%	-10,0%	-3,0%	6,8%	4,8%

Factor Pattern					
	Factor1	Factor2	Factor3	Factor4	Factor5
FI_1018	0.**95991**	-0.00148	-0.02591	-0.00135	0.02946
FI_1019	0.**79963**	-0.00314	-0.03590	-0.00260	0.**49678**
FI_1020	0.**77163**	-0.00310	-0.07015	-0.00480	-0.**43833**
FI_1021	0.08263	-0.00182	0.**99288**	0.08294	-0.02181

Variance Explained by Each Factor				
Factor1	Factor2	Factor3	Factor4	Factor5
3.0610761	2.6409448	1.9972263	1.9840596	1.0315740

Factor Pattern					
	Factor1	Factor2	Factor3	Factor4	Factor5
FI_1018	2,9%	-36,8%	-3,2%	-53,9%	15,3%
FI_1019	2,1%	-18,2%	-0,8%	-44,1%	9,3%
FI_1020	-1,4%	-21,1%	-3,4%	-47,2%	-3,4%
FI_1021	-2,7%	0,0%	0,7%	-40,2%	-1,4%

Variance Explained by Each Factor				
Factor1	Factor2	Factor3	Factor4	Factor5
3.0184591	2.6914142	1.9971643	1.9892570	1.0323149

Table 1 Comparison of different PRIORS option

Table 2 provides different approach for the ROTATE option and even in this case it is clear that the option has minimum effect on the solution. On the other hand, a very interesting result provides comparison between Table 1 a Table 2 – we can see that the rotation of factors helps with interpretability – in the rotated variant factor loadings increased for significant variables of particular factor and decreased with the rest of variables. However, interpretability costs something – ability to explain variance of first and second factor decreased (compare Variance Explained by Each Factor table in Table 1 and Table 2). Nevertheless, in the case of our data set even without rotation data are interpretable and results have economic sense – high factor loadings are between financial indicators in the same group or with similar economic meaning. However, this is a specific situation, thus in general term we would recommend to verify rotation and its benefits for interpretation.

METHOD=P PRIORS=SMC ROTATE=VARIMAX

Rotated Factor Pattern					
	Factor1	Factor2	Factor3	Factor4	Factor5
FI_1001	-0,0283	0,**91226**	-0,03161	0,11798	0,00333
FI_1002	-0,01506	0,04842	-0,00404	0,**79704**	0,02989
FI_1003	0,00623	0,03581	0,**9987**	0,03536	0,00418
FI_1004	0,**76152**	0,02183	-0,00635	0,01166	0,04448
FI_1005	0,**98726**	0,02798	-0,0085	0,01397	0,05809
FI_1006	0,**98692**	0,028	-0,00796	0,01401	0,05806
FI_1008	0,00023	0,00002	-0,00001	0	0,00037
FI_1009	-0,05842	-0,00054	-0,00339	-0,03962	0,**99352**
FI_1011	-0,05832	-0,00067	-0,00256	-0,03825	0,**99358**
FI_1013	0,3168	0,00967	-0,00285	0,00523	0,01902
FI_1015	0,00151	-0,00037	-0,0003	-0,00038	-0,00016
FI_1016	0,00314	0,00558	-0,0086	-0,00214	0,00034
FI_1017	0,00007	-0,00017	-0,00062	-0,0014	0,00047
FI_1018	-0,02833	0,**6745**	0,00461	0,**68306**	0,02497
FI_1019	-0,0222	0,2368	-0,02133	0,**91068**	0,03392
FI_1020	-0,02859	0,**86263**	-0,03511	0,21513	0,00698
FI_1021	0,00623	0,03581	0,**9987**	0,03536	0,00418

Variance Explained by Each Factor				
Factor1	Factor2	Factor3	Factor4	Factor5
2,639013	2,0944358	1,997817	1,997475	1,986139

METHOD=P PRIORS=SMC ROTATE=PROMAX

Rotated Factor Pattern (Standardized Regression Coefficients)					
	Factor1	Factor2	Factor3	Factor4	Factor5
FI_1001	-0,00048	0,**969817**	-0,00687	-0,12075	4,75E-05
FI_1002	-0,0006	-0,14982	-0,00086	0,**856513**	4,49E-05
FI_1003	-0,00025	-0,00356	**1,000000**	-0,00261	-6,13E-06
FI_1004	0,**763244**	-1,9E-05	-0,00046	0,00045	-9E-05
FI_1005	0,**989502**	-8,1E-05	-0,00088	-0,00056	0,000357
FI_1006	0,**989153**	-7,8E-05	-0,00034	-0,00052	0,000347
FI_1008	0,000248	1,33E-05	-1,88E-06	2,81E-06	0,000357
FI_1009	-0,00021	0,00029	-0,00049	-0,00098	0,**99603**
FI_1011	-9,7E-05	-0,00024	0,000339	0,000532	0,**996031**
FI_1013	0,31757	0,00054	-0,00038	0,000467	0,000465
FI_1015	0,001482	-0,00034	-0,0003	-0,00033	-0,00024
FI_1016	0,003347	0,006848	-0,00843	-0,00362	0,00026
FI_1017	8,23E-05	0,000184	-0,00062	-0,00144	0,000524
FI_1018	0,001778	0,**564312**	0,024344	0,**559947**	-4,5E-05
FI_1019	-0,00045	0,028063	-0,01283	0,**929333**	6,24E-05
FI_1020	-0,00054	0,**890859**	-0,01147	-0,001	-3,81E-07
FI_1021	-0,00025	-0,00356	**1,000000**	-0,00261	-6,04E-06

Variance Explained by Each Factor Ignoring Other Factors				
Factor1	Factor2	Factor3	Factor4	Factor5
2,640949	2,565746	2,004882	2,44859	1,984156

Table 2 Comparison of different ROTATE option

3.2 Principal Component Analysis

As mentioned earlier, PCA could be considered as a special case of FA. If we turn off options PRIORS and ROTATE we obtain a Principal Component Analysis result. Nevertheless, SAS disposes with PROC PRINCOMP procedure especially for PCA, but a result is the same with the mentioned variant of PROC FACTOR.

We used the same data set for PCA with results shown below. From Figure 2 it is evident that eigenvalues of first factors are not able to explain that much variability as in the case of FA. The fifth component explains 66% instead of 99% in the case of FA. Within the eighth component PCA is able to explain 84% of variability (still with interpretable factors).

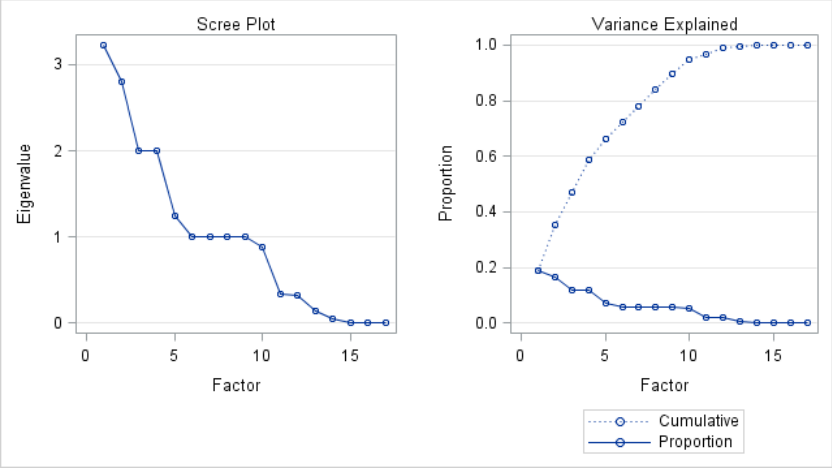


Figure 2 Scree Plot and Variance Explained of PCA

It is very important to decide if our data contains errors. If not and our objective is to explain maximum of original variability then PCA is a good choice, whereas in our case PCA is interpretable and its ability to explain variability with 8 components is more than 84%, hence, our reduction of original variables is more than 50%. On the other hand, if we have doubts about data quality or we rather focus on correlation structures between variables we should use FA with appropriate options, thus we gain correlation structures and more interpretable factors.

4 Conclusion

The article presents ordination analysis with its application in two forms – Common Factor Analysis and Principal Component Analysis. We showed that the type of rotation or definition of prior communality estimates options don't have a big influence on results in FA method, which is usual conclusion in many cases. However, we highlighted an importance of FA and its ability of rotation which helps with interpretation in many economic situations. Both methods are useful techniques which can significantly reduce number of variables and provide correlation structures of variables. PCA is suitable if our goal is a reduction of variables with similar original variability. FA is better for data with errors or for correlation structures and economic interpretation.

Acknowledgements

Supported by a grant No. IGA F4/24/2014, Faculty of Informatics and Statistics, University of Economics, Prague.

We gratefully acknowledge support from Bisnode which provided full data set.

References

[1] Ajamani, V.B.: Applied Econometrics using the SAS System, John Wiley & Sons, Inc., Hoboken, New Jersey, 2009. ISBN: 978-0-470-12949-4

[2] Factor Analysis Using SAS PROC FACTOR. UCLA: Statistical Consulting Group. From http://www.ats.ucla.edu/stat/sas/library/factor_ut.htm. (accessed May 1, 2014)

[3] SAS User's Guide. The Factor Procedure. SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513, USA. From http://support.sas.com/documentation/cdl/en/statug/66103/HTML/default/viewer.htm#statug_factor_overview01.htm (accessed May 1, 2014)

[4] SAS User's Guide. The PRINCOMP Procedure. SAS Institute Inc., SAS Campus Drive, Cary, North North Carolina 27513, USA. From http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_princomp_sect001.htm (accessed May 1, 2014)

Estimation of Technical Efficiency of Food Processing Firms in the Czech Republic over the crisis period

Tamara Rudinskaya

Abstract. The aim of this paper is to estimate technical efficiency of food processing firms in the Czech Republic based on panel data from years 2005–2011 and find out whether crisis, started at the year 2008, had significant impact on technical efficiency. Taking into account the fact, that firm heterogeneity can influence estimation results, we used „True" random effects model without as well as with heteroscedasticity assumption. For empirical study unbalanced panel data set contained 3,889 observations of 839 food processing Czech companies was used. The data set represents 9 branches of food processing industry, 14 regions of the Czech Republic. Using stochastic frontier analysis and z-test we verified that in most cases variables (capital, labour, material and dummy-variable for testing differences in efficiency through the crisis period) have statistically significant impact on the firms' output.

Keywords: Food processing industry, technical efficiency, Stochastic Frontier Analysis.

JEL classification: D24
AMS classification: 90B30

1. Introduction

The crisis of the global economy, which occurred during the years 2008–2011, affected almost all sectors of national economy. Czech industry sector has been facing adverse impacts of the crisis. A number of authors investigated crisis influence to agriculture, manufacture industry and banking sphere. Charoenrat [3] estimated technical efficiency of Thai manufacture in pre- and post-crisis period. For this the author estimated two models (before and after the crisis) using Stochastic Frontier (SFA) and Data Envelopment Approach (DEA). Ersoy [6] used Battese and Coelli [2] model within the framework of SFA for estimation of crisis influence to output results of Turkey banks. In the Czech Republic S. Junková, E. Matušková [10] used Seasonal autoregressive process for the calculation of the likely scenarios for the future development of GDP and its components.

In this paper we analysed productivity of Czech food processing sector through the crisis period. Our main aim is to conduct a comparative analysis among the different periods, i.e. before the crisis (till the year 2008) and after the crisis (after the year 2009) and identify the productive and less productive companies in Czech food processing industry.

An important indicator of productivity of companies is technical efficiency (TE), which is an integral part of overall economic efficiency. Nonparametric approach (DEA), as well as stochastic approach (SFA) has been widely applied. Both approaches assume that firms are not heterogeneous but inefficient, since all inefficiency scores are estimated by assuming a homogeneous technology available to all producers. This suggests that the impact of inefficiency in the agriculture often is overestimated [8]. Considering significant heterogeneity of food processing firms, as was proved in [5], we used „True" random effects (TRE) model for efficiency estimation. These models allow distinguishing time-invariant heterogeneity from time-varying inefficiency.

2 Materials and methods

2.1 Stochastic Frontier Analysis

To study the determinants of TE we used the SFA methodology developed by Aigner et al. [1]. The SFA method is based on an econometric (i.e., parametric) specification of a production frontier. Using a generalized production function and cross-sectional data, this method can be depicted as follows:

$$\boldsymbol{y_i = f(x_{ij}; \beta) * exp(\varepsilon_i)} \tag{1}$$

where $\boldsymbol{y}$ represents output, $\boldsymbol{x}$ is a vector of inputs, $\boldsymbol{\beta}$ is a vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is the error term.

The subscripts $\boldsymbol{i}$ and $\boldsymbol{j}$ denote the firm and inputs, respectively.

In this specific formulation, the error term is farm specific and is composed of two independent components, $\boldsymbol{\varepsilon_i = v_i - u_i}$. The first element, $\boldsymbol{v_i}$ is a random variable reflecting noise and other stochastic shocks entering into

the definition of the frontier, such as weather, luck, strikes, and so on. This term is assumed to be an independent and identically distributed normal random variable with zero mean and constant variance ***iid*** $[\boldsymbol{N}\sim(\boldsymbol{0},\boldsymbol{\sigma_v^2})]$

The second component, $\boldsymbol{u_i}$, captures technical inefficiency relative to the stochastic frontier. The inefficiency term $\boldsymbol{u_i}$ is nonnegative and it is assumed to follow a half-normal distribution [11].
An index for TE can be defined as the ratio of the observed output (***y***) and maximum feasible output (***y****):

$$\boldsymbol{TE_i} = \frac{\boldsymbol{y_i}}{\boldsymbol{y_i^*}} = \frac{\boldsymbol{f(x_{ij};\beta)\cdot exp(v_i - u_i)}}{\boldsymbol{f(x_{ij};\beta)\cdot exp(v_i)}} = \boldsymbol{exp(-u_i)} \tag{2}$$

Because $\boldsymbol{y\leq y^*}$, the TE index is bounded between 0 and 1; TE achieves its upper bound when a firm is producing the maximum output feasible level (i.e., $\boldsymbol{y = y^*}$), given the input quantities. Jondrow et al. [9] demonstrated that farm-level TE for half-normal distribution of inefficiency term can be calculated from the error term $\boldsymbol{\varepsilon_i}$ as the expected value of $-\boldsymbol{u_i}$ conditional on $\boldsymbol{\varepsilon_i}$, which is given by

$$\boldsymbol{E[u_i|\varepsilon_i]} = \frac{\boldsymbol{\sigma_u\sigma_v}}{\boldsymbol{\sigma}}\left[\frac{\boldsymbol{\phi(\varepsilon_i\lambda/\sigma)}}{\boldsymbol{1-\Phi(\varepsilon_i\lambda/\sigma)}} - \frac{\boldsymbol{\varepsilon_i\lambda}}{\boldsymbol{\sigma}}\right] \tag{3a}$$

where $\boldsymbol{\sigma^2 = \sigma_u^2 + \sigma_v^2}$, $\boldsymbol{\lambda = \sigma_u/\sigma_v}$, $\boldsymbol{\phi}(\cdot)$ represent the standard normal density and $\boldsymbol{\Phi}(\cdot)$ the standard normal cumulative density functions. The maximum likelihood estimation of Eq. [3] provides estimators for the variance parameters $\boldsymbol{\sigma_u^2}$ and $\boldsymbol{\sigma_v^2}$.

In the case of exponential distribution of inefficiency farm-level TE is calculated in the form

$$\boldsymbol{E[u_i|\varepsilon_i]} = \boldsymbol{\tilde{\mu}_i} + \boldsymbol{\sigma_v}\left[\frac{\boldsymbol{\phi(-\tilde{\mu}_i/\sigma_v)}}{\boldsymbol{1-\Phi(\tilde{\mu}_i/\sigma_v)}}\right] \tag{3b}$$

where $\boldsymbol{\tilde{\mu} = -\varepsilon - \sigma_v^2/\sigma_u}$

Thus, the TE measure for each farm is equal to $\boldsymbol{TE_i = exp(-E[u_i|\varepsilon_i])}$

"True" random effects model

In the fixed-effects model it is assumed that the inefficiency term is fixed and the correlation with regressors is allowed. Unlike fixed effects model the opposite situation is considered, in which the $\boldsymbol{u_i}$ are randomly distributed with constant mean and variance, but are assumed to be uncorrelated with the regressors and the $\boldsymbol{v_{it}}$. The random effects specification assumes that the firm specific inefficiency is the same every year, i.e. the inefficiency term is time invariant. In this propositions the model absorbs all unmeasured heterogeneity in $\boldsymbol{u_i}$.

Greene [7] argued that the random effects model with the proposed extensions has three significant weaknesses. The first is its implicit assumption that the effects are not correlated with the included variables. The second problem with the random effects is its hypothesis that the inefficiency is the same in every period. For a long time series data, this is likely to be an undesirable assumption. The third shortcoming of this model is that in this model $\boldsymbol{u_i}$ carries both the inefficiency and, in addition, any time invariant firm specific heterogeneity. To avoid the former limitations Greene [7] proposed "True" random effects model that is as follows:

$$\boldsymbol{y_{it}} = \boldsymbol{\alpha} + \boldsymbol{\beta'x_{it}} + \boldsymbol{w_i} + \boldsymbol{v_{it}} \pm \boldsymbol{u_{it}} \tag{6}$$

where $\boldsymbol{w_i}$ is the random firm specific effect and $\boldsymbol{v_{it}}$and $\boldsymbol{u_{it}}$ are the symmetric and one sided components specified earlier.

2.2 Data set

The panel data set was collected from the Albertina database. The database contains all registered companies and organisations in the Czech Republic. The analysis uses information from the final accounts of companies whose main activity is food processing in the period from 2005 till 2011. After the cleaning process (removing companies with missing observations and negative values of the variables), the unbalanced panel data set contains 3,889 observations of 839 food processing Czech companies. The data set represents 9 branches of food processing industry, 14 regions of the Czech Republic.

The following variables were used in the analysis: Output, Labour, Capital and Material. Output is represented by the total sales of goods, products and services of the food processing company. In order to avoid inflation changes, Output was deflated by the price index of food processing companies according to the branch. (2005=100). The Labour input is used in the form of total personnel costs per company, divided by the average annual regional wage. The Capital variable is represented by the value of tangible assets. Material is total costs of material and energy consumption per company. Capital and Material were deflated by the price index of

industrial sector (2005=100). Output, Capital, Material variables are measured in thousand CZK, Labour variable is a coefficient. The share of branches is represented in Table 1.

Year	Average output per firm	Number of observations	Share of branch in industry output[1]						
			CA 101	CA 103	CA 104	CA 105	CA 106	CA 107	CA 109
2005	235031	453	23.3	2.7	5.3	25.8	4.8	8.9	5.4
2006	258171	501	20.3	3.0	3.7	23.7	4.4	13.7	4.6
2007	271466	573	24.1	2.6	4.1	24.8	3.6	12.1	6.5
2008	240675	609	23.5	2.4	2.9	23.2	3.7	12.7	6.7
2009	248213	620	23.2	2.7	1.7	21.1	3.4	11.4	11.8
2010	264734	609	21.5	2.7	1.7	20.6	3.4	13.5	11.9
2011	274232	524	23.6	2.4	1.5	19.4	3.0	11.2	13.8

Source: own processing

Table 1 Significance of main branches in the dataset (as a percent of total industry output)

The most significant share in total output belongs to Preserved meat and meat products (22.8% on average) and Dairy products (22.7% on average). The share of latter is decreasing over the years. The decline might have been caused by the exit of large firms from the branch or the database. Bakery and farinaceous sphere produces on average 11.9% of total industry output. The share of other branches is not very significant. Prepared animal feeds share increased almost twice from the year 2009, whereas producing of oils and fats is decreasing.

For following analysis of technical efficiency 4 branches, namely CA 101, CA 105, CA 107, CA 109, were chosen. These branches have the most significant share in total industry output and mostly represent it. Remain branches will be missed because of the small sample size and insignificance of output results.

3 Empirical application

The empirical analysis is based on the estimation of a Cobb-Douglas production function in which both the output and inputs are expressed in logarithmic form. Hence, the estimated coefficients reflect the output elasticities [11]. The empirical stochastic frontier production model that was used is specified as follows:

$$\boldsymbol{ln\, Y_{it} = \beta_0 + \beta_K\, ln\, K_{it} + \beta_L\, ln\, L_{it} + \beta_M\, ln\, M_{it} + v_{it} - u_{it}} \quad (7)$$

where $\boldsymbol{\beta_0, \beta_K, \beta_L, \beta_M}$ are parameters of the model, $\boldsymbol{K_{it}}$ - Capital variable, $\boldsymbol{L_{it}}$ - Labour variable, $\boldsymbol{M_{it}}$ - Material variable, $\boldsymbol{v_{it}}$ - error term (normal distribution), $\boldsymbol{u_{it}}$ - inefficiency term. To calculate the inefficiency of particular farm the Jondrow et al. [9] estimator was used.

There is plenty of empirical evidence of distributional sensitivity of $\boldsymbol{v_i}$. However, what is not so clear is whether an inefficiency term is sensitive to distribution assumption [11]. Ritter and Simar argue for use of a relatively simple distribution such as half-normal or exponential [12]. Follow this, for the purpose of empirical research we assume the inefficiency distribution to be half-normal $\boldsymbol{u_i \sim N^+(0; \sigma_u^2)}$ and exponential $\mathbf{u_i \sim \varepsilon(\sigma_u)}$. These distributions are single-parameter distributions with zero mean.

We introduce dummy variable to find out whether efficiency of food processing firms changed through the crisis period.

The hypothesis that heteroscedasticity is no present in error term was rejected according to the White and Breusch-Pagan test. Disregarding the heteroscedasticity has undesirable consequences [11]

- Ignoring the heteroscedasticity of $\boldsymbol{v_i}$ still gives consistent estimations of the frontier function parameters except for the intercept term which is biased. Estimates of the technical efficiency are biased.
- Ignoring the heteroscedasticity of $\boldsymbol{u_i}$ causes biased estimations of the frontier function parameters as well as estimates of technical efficiency.
  In this connection we estimated and compared models with and without heteroscedasticity.

The maximum likelihood Estimates (MLE) of the stochastic production parameters are presented in Table 2. For estimation software Stata 11.2 was used.

[1] CA 101- Preserved meat and meat products, CA 103- Processed and preserved fruit and vegetables, CA 104- Vegetable and animal oils and fats, CA 105- Dairy products, CA 106- Grain mill products, starches and starch products, CA 107- Bakery and farinaceous products, CA 109- Prepared animal feeds

Branch	Parameters	Exponential distribution		Half-normal distribution	
		homoscedasticity	heteroscedasticity	homoscedasticity	heteroscedasticity
CA 101	$\beta_k (\ln K_{it})$	.0443845	.0539977***	.213794	.0350038
	$\beta_L (\ln L_{it})$	.3216866***	.1985362***	.2899398***	.1261898***
	$\beta_M (\ln M_{it})$	.6573683***	.724475 ***	.5182894***	.710005***
	dummy	-.0254908***	-.0447764***	-.0442224***	-.0534125***
	Intercept	4.087348	3.321423***	4.2542	3.899763
	λ	2062412	2.8025864	303673.1	4.8618997
CA 105	$\beta_k (\ln K_{it})$	.0270569*	.0435546***	.0242648	.0475363***
	$\beta_L (\ln L_{it})$	.1473224***	.1324345***	.143774***	.1311356***
	$\beta_M (\ln M_{it})$	.7585351***	.7681266***	.7616235***	.7824753***
	dummy	-.06647***	-.075167***	-.0719828***	-.0799192***
	Intercept	3.211201***	2.906244***	3.23216***	2.722446***
	λ	1.76665	1.330602	3.751627	2.132711
CA 107	$\beta_k (\ln K_{it})$	.0252472	.0271481***	.1067544	.0422329
	$\beta_L (\ln L_{it})$	.1349695***	.1917291***	.1783642***	.0943256***
	$\beta_M (\ln M_{it})$	.483673***	.7286962***	.5524091***	.7421753***
	dummy	.0198949***	.0043961***	.0302282***	.0131896***
	Intercept	5.321197	3.404458***	4.157057	3.094514
	λ	2674661	1.729769	1558777	3.851412
CA 109	$\beta_k (\ln K_{it})$	.1021859	.0269374*	.0286194***	.0152294
	$\beta_L (\ln L_{it})$	.2140034***	.1320802***	.2501556***	.1504309***
	$\beta_M (\ln M_{it})$	.6205995***	.7813021***	.6969415	.7546808***
	dummy	-.0300814	.0206327	-.0614259***	.0045324
	Intercept	4.060983	2.751047***	4.051301	3.246163
	λ	4611703	1.788370	491997.2	3.238670

Source: own processing; ***, **, * denotes significance at the 1%, 5%, and 10% level, respectively

Table 2 Estimated stochastic frontier

CA 101: Estimated parameters Labour, Material and dummy variable are significant under z-test in the case of all models. Heteroscedastic model with exponential distribution of inefficiency term shows significance of all parameters estimated. Signs of the coefficients are found to be consistent with economic theory. In Cobb-Douglas model these coefficients denotes the variation or possible percentage change in aggregate output as a result of one percent change in the input, that is production elasticity. All production elasticities are positive; the highest elasticity displays production factor Material. The parameter $\boldsymbol{\lambda}$ is the relation between the variance of $\boldsymbol{u_{it}}$ and $\boldsymbol{v_{it}}$. Thus, the parameter indicates the significance of technical inefficiency in the residual variation. A value larger than one suggests that variation in $\boldsymbol{u_{it}}$ prevails the variation in the random component $\boldsymbol{v_{it}}$. Dummy variable is significant at the 5% significance level in all the cases; the sign is negative, what supposes the higher output at the period after the year 2008.

CA 105: The parameters of the model are statistically significant at the 1% level, except Capital, that exhibits low significance in homoscedastic Normal-Exponential model and insignificance in homoscedastic Normal-Half-Normal model, the slopes of coefficients are positive, that is consistent with economic theory. Dummy variables display lower output at the period before crisis. The parameter $\boldsymbol{\lambda}$ is more than one indicates presence of inefficiency in the data.

CA 107: The criteria of theoretical consistency, i.e., the assumptions regarding slope of the production function, are fulfilled in the case of all models. Elasticity of the production factor capital is low; the parameter is significant only in the case of Normal-Exponential model with heteroscedasticity. Dummy variable is significant, has positive sign, and indicates that firm had higher output before the crisis period started. The parameter $\boldsymbol{\lambda}$ indicates the significance of inefficiency in the residual variation.

CA 109: Signs of the coefficients, as well as the numerical results obtained, were found to be robust under the Normal-Exponential model with heteroscedasticity. The parameter $\boldsymbol{\lambda}$ suggests that efficiency differences among firms are an important reason for variations in production. Dummy variable is significant in the case of homoscedastic model with half-normal distribution of inefficiency and insignificant in the other cases, what denies the considerable change of efficiency through the crisis period.

The results of parameters estimation with half-normal and exponential distribution of inefficiency are quite similar with moderate difference.

The results of technical efficiency estimation and returns to scale calculation are represented in Table 3. Since Cobb-Douglas model for empirical analysis was used, estimated parameters of variables represent elasticity. Returns to scale in this case can be calculated by summing of elasticities of estimated factors.

Branch		Exponential distribution		Half-normal distribution	
		homoscedastic	heteroscedastic	homoscedastic	heteroscedastic
Technical efficiency					
CA101	Mean	.6256706	.7241466	.4587574	.5834264
	Median	.7010763	.7969072	.4886918	.6524997
CA105	Mean	.8945684	.9086161	.8594351	.8873397
	Median	.9309236	.9341818	.8916587	.9060127
CA107	Mean	.5971694	.7478524	.5108505	.6707211
	Median	.669158	.8185582	.5423648	.736432
CA109	Mean	.7083643	.8206289	.6074582	.7524448
	Median	.7840143	.8665592	.6542178	.8010049
Returns to scale					
CA101		1.0234394	.9770089	1.0220232	.8711986
CA105		.9329144	.9441157	.9296623	.9611472
CA107		.6438897	.9475734	.8375277	.8787338
CA109		.9367888	.9403197	.9757165	.9203411

Source: own processing

Table 3 Technical efficiency and returns to scale

The models with exponential distribution of inefficiency term exhibit higher efficiency than the model with half-normal distribution by 14.8% on average under the assumption of homoscedasticity and by 10.1% under the heteroscedasticity assumption. The models under the assumption of heteroscedasticity in error and inefficiency term with different distribution demonstrate higher efficiency estimation (by 14.6% in Normal-Exponential and 16.9% in Normal-Half-Normal model) than the model under homoscedasticity assumption. Comparing efficiency among the branches showed that CA105-Dairy products has the highest efficiency (88.7% on average), the lowest belongs to CA101- Meat products (59.8% on average).

CA 101: the highest level of efficiency was got under the assumption of exponential distribution of inefficiency term and heteroscedasticity of error and inefficiency term (72.4%), 50% processing firms in this branch are operating under 79.7% level of efficiency, or about 20.3% of the potential output is lost due to technical inefficiency. The branch is characterised by diminishing returns to scale in the case of heteroscedastic half-normal model and almost constant returns to scale in other cases.

CA 105: mean efficiency according to the model under heteroscedasticity assumption with exponential distribution is 90.9%, the median value is 93.4%, the highest level among the exploring branches. The branch exhibits slightly decreasing returns to scale.

CA 107: the highest difference between the model under homoscedastic and heteroscedastic assumptions belongs to this branch (the average difference is 24.5%). Returns to scale indicator shows large variance depending on the assumption, demonstrating averagely diminishing returns to scale.

CA 109: the highest level of efficiency demonstrates the model under heteroscedasticity assumption and exponential distribution of inefficiency term (82.1%). In this case 50% processing firms produce 86.7% of the maximum output. All models demonstrate slightly decreasing returns to scale.

Central to the stochastic frontier model is the one-sided error specification which represents technical inefficiency. It is therefore important to test the existence of the one-sided error in the model. If evidence for the one-sided error specification is not found, the model then reduces to a standard regression model for which a simple OLS estimation would suffice. This amounts to a test for the presence of $\boldsymbol{u_i}$ in the model, and a generalized likelihood ratio (LR) test of the null hypotheses of one-sided error can be constructed based on the log-likelihood values of the OLS (restricted) and the SF (unrestricted) model [11]. Coelli, Rao et.al [4] showed that the test has a mixture of chi-square distribution. The result of LR test of our empirical models indicates a rejection of the null hypotheses of no technical efficiency in both cases.

The LR test can be used for testing hypothesis concerning the parameters of the models, including distribution assumptions. In our instance this test is not suitable, because half-normal and exponential distribution, used for technical efficiency estimation, have the same number of parameters, and the number of degrees of freedom is equal to zero. For alternative model selection, Akaike (AIC) and Bayesian Information

Criterion (BIC) can be used. The results show that the production model under the assumption of exponential distribution of inefficiency term has lower AIC and BIC criterion, therefore fits the data better comparing to the half-normal model.

4. Conclusion

The aim of the paper was technical efficiency estimation of food processing Czech firms through the crisis period. The study, based on SFA and estimation of the TRE model under the assumption of exponential and half-normal distribution of inefficiency term, showed that the highest efficiency among the processing industries in the Czech Republic has CA105- Dairy products.

The technical efficiency indicator estimated using TRE model under homoscedasticity and heteroscedasticity assumption with different distribution of inefficiency term, exhibits different results. The models with exponential distribution showed, on average, by 14.8% higher efficiency than the model with half-normal distribution under the assumption of homoscedasticity and by 10.1% under the heteroscedasticity assumption. The models under the assumption of heteroscedasticity in error and inefficiency term with different distribution demonstrate higher efficiency estimation (by 14.6% in Normal-Exponential and 16.9% in Normal-Half-Normal model) than the model under homoscedasticity assumption.

The study showed that differences in efficiency before and after 2008 year exist; moreover firms before the crisis period are less efficient. The average efficiency level of processing firms in the sample was 59.8% (CA101), 88.7% (CA105), 63.2% (CA107), and 72.2% (CA109). This result implies that farmers were able to obtain indicated potential output from a given combination of production inputs. All branches exhibit constant or slightly diminishing returns to scale, except CA107- Bakery and farinaceous products, which has rather diminishing returns to scale. The average level of TE in the sample suggests that, from a technical standpoint, the opportunity exists to expand production output using the current level of inputs and the technologies already available in the firm. These results suggest that food processing companies can improve their productivity and efficiency if they take advantage of more efficient firm practices.

Acknowledgements

The research is financed from grant No.11110/1312/3143„Inter- a intra-sektorální změny efektivnosti a produktivity v českém potravinářském průmyslu- systematický nebo idiosynkratický vývoj?“ of the IGA, CULS.

References

[1] Aigner, D., Lovell, C.A., Knox, Schmidt, P.: Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* **6** (1977), 21-37

[2] Battese, G.E., Coelli, T.J.: Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis* **1** (1992), 153-169, ISSN 0895-562X

[3] Charoenrat, T.: The technical efficiency of Thai manufacturing small and medium sized enterprises: a comparison between the pre- and post-financial crisis. *Research online (2012).* Available at: http://ro.uow.edu.au/cgi/ viewcontent.cgi?article=4768&context=theses

[4] Coelli Timothy, J., Prasada Rao, D.S. et al.: *An Introduction to Efficiency and Productivity Analysis*. Second edition. New York: Springer, 1995, 349p., ISBN 978-0387-24266-8

[5] Čechura, L.: *Zdroje a limity růstu agrárního sektoru. Praha*: Wolters Kluwer, 2009, 296p., ISBN 978-80-7357-493-2

[6] Ersoy, I.: *The Impact of the Global Financial Crisis on the Efficiency of Foreign Banks in Turkey*, 2009 [Online]. Available at: http://www.opf.slu.cz/kfi/icfb/proc2009/pdf/08_Ersoy1.pdf

[7] Greene, W.: Fixed and Random Effects in Stochastic Frontier Models. *Journal of Productivity Analysis* **23** (2005), 7–32

[8] Hockmann H., Pieniadz A.: Farm Heterogeneity and Efficiency in Polish Agriculture: A Stochastic Frontier Analysis. 12th Congress of the European Association of Agricultural Economists – EAAE 2008. [Online]. Available at: http://ageconsearch.umn.edu/bitstream/44278/2/496.pdf

[9] Jondrow, J., Lovell, C. A. K, Materov, I. S., Schmidt, P.: On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. *Journal of Econometrics* **19** (1982), 233-238, ISSN 0304-4076

[10] Junková, S., Matušková, E.: The influence of crisis on the sector structure of economy focusing on agriculture. *Agris on-line Papers in Economics and Informatics* **2** (2011). [Online]. Available at: http://ageconsearch.umn. edu/bitstream/109735/2/agris_online_2011_2_junkova_matuskova.pdf

[11] Kumbhakar, S.C., Lovell, C.A. *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge, 2000

[12] Ritter, C., Simar, L. Pitfalls of Normal-Gamma Stochastic Frontier Models. Journal of Productivity Analysis **8:2** (1997)

Conditional Volatility Model Selection and Comparison: Example from Stock Markets

Petr Seďa[1]

Abstract. Model selection and model comparison of the second moment of a given random variable has been widely discussed topic in financial econometrics for many years. In this paper, we focus on model comparison and model selection taking in consideration univariate conditional volatility models. The aim of this paper is to provide comparison of linear and nonlinear ARCH family models based on out-of-sample forecast using two different approaches. In first approach, alternative volatility models are compared by different loss functions based directly on variance forecast and respective Diebold-Mariano type tests. The comparison of models based on Diebold-Mariano test is limited since it allows pairwise comparison only. Therefore, the model confidence set (MCS) technique will be utilized as well. The second approach is relatively new and includes indirect evaluation methods which consider using of alternative variance forecasts, in particular a VaR framework. In order to illustrate application of both mentioned approaches, we consider daily data of WIG20 and PX indexes in the period of 2004-2012 years which also includes a period of recent global financial crisis of 2008-2009 years.

Keywords: GARCH, MCS approach, out-of-sample forecast, TGARCH, VaR.

JEL Classification: C52, C53, C58
AMS Classification: 62P20

1 Introduction

Financial markets, due to their key role in the economics, have been studied from different points of view. In this regard, one key aspect of the financial markets that has attracted much attention in financial econometrics is the analysis of respective returns and its volatility. Volatility is usually a symptom of a highly liquid financial market. Investors may interpret a growth in financial market volatility as an increase of the risk of investment and consequently they may shift their funds to less risky assets.

Model comparison and model selection especially in case of conditional volatility has been widely and often discussed in financial sciences for an extended period. As a final output, several approaches have been developed in order to evaluate if a given model is able to capture empirical patterns observed on sample data set. The other problem is to check if there exists a preference in favor of any alternative model that might be given for purposes of analysis provided. In this paper, we focus on model comparison and model selection in a specific framework, namely univariate conditional volatility models of high frequency financial time series.

In the financial econometrics literature many different techniques have been utilized to model volatility. The autoregressive conditional heteroskedasticity (ARCH) model derivate by [5], and later generalized by [1] is one of the most popular methods used for modeling high-frequency return series in financial engineering. While volatility in the standard GARCH-type model responds equally to positive and negative return shocks, asymmetric GARCH models allow positive and negative shocks to have different impacts on future volatility values. To model this phenomenon in this study it will be applied model that allows analyzing the asymmetric shocks to volatility, namely the Threshold ARCH model.

The aim of this paper is to provide comparison of selected linear and nonlinear conditional volatility models based on out-of-sample forecast by two different approaches using sample data from Central European stock market. Namely, we consider daily data of WIG20 and PX indexes in the period of 2004-2012 years. First analytical approach is based on idea that alternative models are contrasted by different loss functions that are based directly on variance forecast and respective Diebold-Mariano type tests [4]. Second technique includes indirect evaluation methods which consider the model confidence set (MCS) framework and using of alternative variance forecasts that is based on Value at risk [2].

The reminder of this paper is as follows. First, the methods of model selection and comparison that consider direct and indirect model evaluation based on out-of-sample forecast will be described shortly. Second, the linear

[1] VŠB-Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská třída 33, 701 21 Ostrava 1, petr.seda@vsb.cz .

and nonlinear conditional volatility models will be introduced. In Section 4, the data sample will be described and briefly statistically analyzed. Section 5 reports empirical findings as provided on Central European stock market data sample, and Section 6 concludes our analysis.

2 Approaches of model selection and evaluation

The aim of this chapter is to describe some approaches used in this paper as a tool for univariate conditional volatility model comparison and model selection. Selection of the best suitable model might be usually based on in-sample or/and out-of-sample criteria. In this paper, we get focused just on out-of-sample selection criteria. Nevertheless, a practical identification of an optimal volatility model requires an optimal balance between these two groups of criteria. In empirical studies, a trade-off between these two approaches actually exists [3]. A structural analysis of a model estimated will have a greater focus on in-sample outcomes, while a forecasting will have a greater emphasize on out-of-sample fit. The out-of-sample criteria in case of linear and nonlinear ARCH models may be considered from two different perspectives. First approach is represented by a direct evaluation and comparison of volatility forecast, while an indirect comparison of volatility constitutes the second approach.

2.1 Methods of direct model evaluation

Direct model evaluation and comparison technique is based on assumption that some alternative models are usually compared by statistical tests which are directly based on variance forecast. Let define $\hat{\sigma}_t^2$ as variance forecast of a model at time *t* and σ_t^2 as unknown variance at time *t*. For each estimated model one can evaluate a set of standard loss functions like mean absolute error (MAE), mean squared error (MSE) or quasi likelihood (QLIKE). Model equivalence can be verified by more formally using the approach defined by [4]. This approach compares alternative models using loss functions differentials. As proved in [7], although many loss functions exist, the MSE and QLIKE are loss functions that belong to a family of loss functions are robust to a noise in the volatility proxy. Therefore, in this paper the QLIKE loss function will be used as given below:

$$QLIKE = \frac{1}{T}\sum_{t=1}^{T}\left(\ln\hat{\sigma}_t^2 + \frac{s_t^2}{\hat{\sigma}_t^2}\right)^2, \tag{1}$$

where s_t^2 is a proxy for unobserved volatility σ_t^2.

Alternative volatility models can be compared by testing the predictive ability assuming the null hypothesis of the unconditional expected null loss function difference as given below:

$$H_0 : E\left[QLIKE(j)\right] - E\left[QLIKE(i)\right] = E\left[QLIKE(j) - QLIKE(i)\right] = 0. \tag{2}$$

Diebold and Mariano [4] defined under mild conditions of the loss function differences the test statistic τ_{QLIKE} that takes the form:

$$\tau_{QLIKE} = \frac{\overline{lf}_{QLIKE}(j,i)}{\sqrt{Var\left[lf_{QLIKE,t}(j,i)\right]}}, \tag{3}$$

where $Var\left[lf_{QLIKE,t}(j,i)\right]$ denotes a variance estimator which is heteroskedasticity and autocorrelation consistent and $\overline{lf}_{QLIKE}(j,i) = \frac{1}{T}\sum_{t=1}^{T}\left[\left(s_t^2 - \hat{\sigma}_{j,t}^2\right)^2 - \left(s_t^2 - \hat{\sigma}_{i,t}^2\right)^2\right]$.

A comparison based on Diebold-Mariano test represents pairwise comparison only, so that it is not possible to exclude in advance the possibility of having different model rankings linked with different loss functions. Recent literature suggests several approaches that have attempted to resolve this problem. In this paper, the model confidence set as proposed by [6] will be used. The MCS approach constitutes a testing framework for the null hypothesis of equivalence across models. Let define the *M* as a set of different models which have been used for forecasting. The null hypothesis of the MCS is defined as follows:

$$H_0 : E\left[lf_{QLIKE,t}(j,i)\right] = 0, i > j, \forall i, j \in M. \tag{4}$$

The null hypothesis can be verified using the following two test statistics:

$$t_R = \max_{i,j \in M} \left| \frac{\overline{lf}_{QLIKE}(j,i)}{\sqrt{Var\left[lf_{QLIKE,t}(j,i)\right]}} \right|, \tag{5}$$

$$t_{SQ} = \sum_{i,j \in M, j>i} \left(\frac{\overline{lf}_{QLIKE}(j,i)}{\sqrt{Var\left[lf_{QLIKE,t}(j,i)\right]}} \right)^2. \tag{6}$$

The both test statistics are based on bootstrap estimation of variance $Var\left[lf_{QLIKE,t}(j,i)\right]$. More detailed information can be found in [6].

2.2 Methods of indirect model evaluation

Indirect evaluation methods consider using of alternative variance forecasts. For example, conditional variances might be used to price derivatives or to define the market risk exposure of a portfolio. Many studies has been recently addressed to this topic, is focused especially on multivariate models. For instance, Caporin [2] dealt with the evaluation of alternative GARCH models specifications within a Value at Risk (VaR) framework. These methods have focused on the evaluation of alternative VaR forecasts. These forecasts used to determine if a model is more appropriate or not with respect to competitors in determining the future expected risk of a financial instrument. Within this framework, we consider a variable x_t characterized by a time-varying mean and with an unspecified conditional density and showing heteroskedasticity as referred below:

$$x_t \mid I^{t-1} \sim f\left[x_t, \mu_t, \sigma_t^2, \theta\right]. \tag{7}$$

where I^{t-1} denotes the information set up to time *t*-1 and θ is a vector containing additional parameters. Let define the one day VaR for x_t as given below:

$$\alpha = \int_{-\infty}^{VaR(x_{t+1},\alpha)} f\left(x_{t+1}, E\left[\mu_{t+1} \mid I^t\right], E\left[\sigma_t^2 \mid I^t\right], \hat{\theta}\right) dx_{t+1}, \tag{8}$$

where time varying mean and variance are substituted by their conditional expectations, and α represents the VaR confidence level.

The evaluation of alternative variance specifications with a help of VaR could follow two approaches as proved in [2]. In this paper, we provide an interpretation of VaR model comparisons with a help of the MCS approach which would seem to be novel. According [2] we suggest the following loss functions based on VaR forecast:

$$\mathrm{IF} = I\left(x_t < \mathrm{VaR}\left(x_t, \alpha\right)\right), \tag{9}$$

$$\mathrm{PIF}_t = I\left(x_t < \mathrm{VaR}\left(x_t, \alpha\right)\right)\left(1 + \left(x_t - \mathrm{VaR}\left(x_t, \alpha\right)\right)^2\right). \tag{10}$$

The indicator loss function IF identifies exceptions, while the penalized indicator function PIF penalizes exceptions by means of squared deviation between respective returns and VaR. The main benefit of these loss functions lies in a fact that they don´t rest on the true volatility but depend on the volatility forecast. These methods can be therefore used within the MCS approach to compare alternative models.

3 Linear and nonlinear conditional volatility models

In this section, we illustrate well-known univariate volatility models. We consider traditional symmetric GARCH(1,1) model and its extension, the TGARCH (1,1) model, which is able to capture asymmetry in variance. In empirical applications, we search for models that capture the features of analyzed data that provide accurate out-of-sample forecasts. Normally, the GARCH(1,1) model taking the form of discrete data and considering heteroskedasticity may be according to [1] defined as follows:

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \tag{11}$$

where σ_t^2 is conditional variance estimated here and ε_{t-1}^2 denotes news about volatility from the previous period. β_0, β_1, β_2 are constants satisfying the conditions of $\beta_0 > 0$, $\beta_1 + \beta_2 < 1$.

The asymmetric response of good and bad news to future volatility, or the leverage effect, is such that bad news should increase future volatility while good news should decrease future volatility. An extension of the classic GARCH model that allows for leverage effects is the Threshold-ARCH or TGARCH model. The idea of TGARCH model is to divide the distribution of the innovations into disjunctive intervals and then approximate a piecewise linear function for the conditional standard deviation or the conditional variance respectively. Original GARCH model was extended in [9] by including the lagged dummy variable S_{t-i} as a regressor, which is known as the TGARCH model. TGARCH(*p*,*q*) model is therefore estimated with the following equation:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{p} \gamma_i S_{t-i} \varepsilon_{t-i}^2, \tag{12}$$

where

$$S_{t-i} = \begin{cases} 1 \text{ if } \varepsilon_{\text{t-i}} < 0 \\ 0 \text{ if } \varepsilon_{\text{t-i}} \geq 0. \end{cases} \tag{13}$$

In other words, depending on the γ_i being above or under the threshold value which equals zero, ε_{t-i}^2 will have different effects on the conditional variance σ_t^2. When ε_{t-i} is positive a total effect is given by $\alpha_i \varepsilon_{t-i}^2$; while when ε_{t-i} is negative a total effect is given by $(\alpha_i + \gamma_i)\varepsilon_{t-i}^2$. This is why in the case of TGARCH we expect γ_i to be positive, so that bad news would have a more powerful effect on volatility than good news. To accept the null hypothesis of no leverage effect in TGARCH model, the γ_i coefficient must be negative. In other words, if the γ_i coefficient is not negative there is evidence of leverage effects in the series.

4 Data sample description

The original data set used in this paper consists of PX and WIG20 indexes. Empirical analysis is performed on daily data in period from 2004 till 2012, which includes total of 2135 observations. This period was chosen purposely, to compare changes of the Central European equity markets volatility forecast during time with a special emphasis on the behaviour in the time before, during and after the global financial crisis of 2008-2009 years, and also taking into consideration local economic and political events. As it has been already empirically confirmed in [8], crises are not devoted to developed markets only. We have more than 8 years long time series of the closing prices of PX and WIG20 indexes. We divided the basic period into three testing periods. First, the pre-crisis period, was defined from January 2004 to the end of June 2007 (922 observations), second, the crisis period, started at the beginning of July 2007 and finished by March 2009 (429 observations) and the post-crisis period was defined from April 2009 to the middle of March 2012 (784 observations). Our goal is to investigate and compare quality of model selection and evaluation in all three periods. The returns r_t at time *t* were defined as the logarithm of respective indices *p*, that is, $r_t = \ln(p_t - p_{t-1})$.

	Period	Mean	Standard deviation	Skewness	Kurtosis	*J-B* test
PX	pre-crisis	0.0011	0.0107	-0.5572	8.6038	1254.09*
	crisis	-0.0025	0.0251	-0.3939	12.1045	1492.78*
	post-crisis	0.0006	0.0152	-0.0439	5.7086	239.91*
WIG20	pre-crisis	0.0009	0.0127	-0.3235	4.5116	103.87*
	crisis	-0.0022	0.0218	-0.2187	4.7066	55.48
	post-crisis	0.0006	0.0158	-0.0234	5.4143	190.48*

Table 1 Descriptive statistics of PX and WIG20 returns

Table 1 shows several descriptive statistics and the results of the unit root test for both Central European stock returns. Symbol * means a rejection of relevant null hypothesis at 5% significance level. The means of both sample returns are quite small with negative values in crisis period, while the standard deviations are significantly higher, especially in crisis period. Based on the values of skewness, increased values of kurtosis and *J-B* test of normality, the daily PX and WIG20 return series show mostly leptokurtic distribution which has a higher peak and heavy tail, instead of normal distribution. The hypothesis of normality was rejected in all periods on both markets.

5 Empirical findings

In this section, estimation results will be presented. The aim of this section is to demonstrate and present an empirical comparison of the methods discussed and described in the previous sections. The studies will be summarized as described in the Introduction section. For both return series we estimated two specific conditional heteroskedasticity models, namely symmetric linear GARCH(1,1) and nonlinear TGARCH(1,1) models. Both models are estimated on a rolling basis, using a window of 850 (in pre-crisis period), respectively 350 (in crisis period) and 700 (in post-crisis period) observations, under assumption of GED distribution of errors, and then are used to estimate one-day-ahead volatility forecast. The models are compared using the methods described in Sections 2 and 3. In particular, we consider the Diebold-Mariano test using the QLIKE loss function across the models, the MCS approach using again the QLIKE loss function, and the loss functions defined by (9) and (10) at 5% VaR level.

The out-of-sample comparisons start from the outcomes of the Diebold-Mariano test (direct model comparison) using the QLIKE loss function. Our aim is to evaluate and compare GARCH and TGARCH models performance across different market phases. Therefore, we consider separately pre-crisis, crisis and post-crisis periods. To sum up, we compare different data sample with very different volatility and returns. Table 2 reports some characteristic empirical findings. Diebold-Mariano test evaluates the null hypothesis of zero expected difference between loss functions of GARCH and TGARCH models. Significant values at 5% confidence level indicate a preference of linear GARCH model if the value of Diebold-Mariano test is positive (in italics) or asymmetric TGARCH model when the value of Diebold-Mariano test statistic is on the other hand negative (in bold).

Diebold-Mariano test	QLIKE (pre-crisis)	QLIKE (crisis)	QLIKE (post-crisis)
PX	0,245	**-2,692**	**-2,258**
WIG20	-0,627	**-2,273**	-1,864

Table 2 Diebold-Mariano test statistics

Focusing on QLIKE loss functions, both empirical models seem very similar for both stock market indices. The null hypothesis of zero loss function differentials is not rejected only in pre-crisis period, while in crisis and post-crisis periods (only in case of PX) the value of test statistic is negative and statistically significant at 5% significance level. So, there is a clear preference for nonlinear TGARCH model in both Central European stock markets. However, using the other loss functions can naturally lead to different results. Though it is not a case of this study, the limitation of the Diebold-Mariano test lies in possibility to consider just pairwise comparison. As suggested in Section 2.1, the MCS approach overcomes this restrictive comparison. The MCS results based on the t_R statistics for all three testing periods are reported in Table 3. For both return series we evaluated two alternative volatility models using the QLIKE function as well as the IF and PIF loss functions as defined in (9) and (10). To sum up, Table 3 shows the model confidence set over different loss functions, return series and testing periods. Bold values denote that models are included at 5% confidence interval in the confidence set.

If we consider PX return series, the results differs between models and out-of-sample testing periods. In pre-crisis period, QLIKE doesn´t include volatility models into confidence set at all, while IF and PIF indicate again rather preference for linear GARCH. In crisis period, some differences appear across the loss functions and models. According QLIKE, the optimal model is clearly TGARCH. IF excludes GARCH model from the confidence set, while PIF considers both models equivalently. In post-crisis period, there is also preference for nonlinear model according QLIKE function, while IF and PIF consider again both models equivalently. Very similar patterns are observed for Polish stock market index represented by WIG20 index.

To summarize, our findings may be interpreted as confirmation of the direct model comparison outcomes as reported in Table 2 that indicates a preference rather for nonlinear TGARCH model in periods of instability. The former indicate a preference for GARCH, while the latter tent to support TGARCH. When we give a large relevance to volatility spikes, most models appear relevant. However, model preference depends on loss function and on the sample period used for model evaluation.

		QLIKE loss function	IF (5%)	PIF (5%)
PX	GARCH (pre-crisis)	0,07	**0,62**	**0,78**
	TGARCH (pre-crisis)	0,08	0,07	0,01
	GARCH (crisis)	0,04	0,04	**0,41**
	TGARCH (crisis)	**1,00**	**0,22**	**1,00**
	GARCH (post-crisis)	0,05	**1,00**	**1,00**
	TGARCH (post-crisis)	**1,00**	**0,34**	**0,57**
WIG20	GARCH (pre-crisis)	**0,10**	**1,00**	**1,00**
	TGARCH (pre-crisis)	**0,54**	0,05	0,02
	GARCH (crisis)	0,08	**0,39**	**0,38**
	TGARCH (crisis)	**0,54**	**1,00**	**1,00**
	GARCH (post-crisis)	0,01	0,07	**0,13**
	TGARCH (post-crisis)	**0,78**	**1,00**	**1,00**

Table 3 Model confidence set

6 Conclusions

In this paper, we applied on some recently developed methods for univariate conditional volatility model selection and model comparison. We consider selected relatively new approaches which are based on evaluation of out-of-sample forecast of conditional volatility models. In first approach, alternative models were contrasted by QLIKE loss function based directly on variance forecast and Diebold-Mariano type tests. In order to solve limitation of this method we also considered the MCS approach which is able to compare more than two models. The second approach included indirect evaluation methods which consider using of alternative variance forecasts. Within this approach, a model evaluation consider more general loss functions based on VaR forecasts that are compared using the MCS approach. In our study, we dealt with the evaluation of linear GARCH and nonlinear TGARCH models within a VaR framework.

Empirical comparison of the methods discussed above was demonstrated on illustrative example using sample data from Czech and Polish stock markets. We considered daily data of PX and WIG20 indexes in the period of 2004 - 2012 years which includes the stage of recent global financial crisis of 2008-2009 years. Sample results suggest that there is a strong preference for nonlinear models in both stock markets in periods of instability. However, a model preference also depends on the sample period used for evaluation and on the loss function we considered. Both approaches can be simply used on the forecasts produced by other univariate models as well as multivariate specifications. The use of these techniques can be extended to cover also other volatility models as well other loss functions based on VaR forecasts.

Acknowledgements

This paper was supported by the Czech Science Foundation through project No. 13-13142S and the European Social Fund within the project CZ.1.07/2.3.00/20.0296.

References

[1] Bollerslev, T.: Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31** (1986), 307-327.

[2] Caporin, M.: Evaluating Value-at-Risk measures in presence of long memory conditional volatility. *Journal of risk*, **10** (2008), 79-110.

[3] Caporin, M. and McAleer, M. J.: *Model Selection and Testing of Conditional and Stochastic Volatility Models*. Econometric Institute Research Papers EI 2010-57, Erasmus University Rotterdam, Erasmus School of Economics, Econometric Institute, 2010.

[4] Diebold, F. X. and Mariano, R. S.: Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13** (1995), 253-263.

[5] Engle, R. F.: Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50** (1982), 987-1007.

[6] Hansen, P. R, Lunde, A. and Nason, J. M.: The model confidence set. *Econometrica* **79** (2011), 453-497.

[7] Patton, A. J.: Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* **160** (2011), 246-256.

[8] Seďa, P.: Impact of the Global Financial Crisis on Stock Market Volatility: Evidence from Central European Stock Market. In: *Mathematical Methods in Economics 2012*. Karviná, 2012, 787-792.

[9] Zakoian, J. M.: Threshold Heteroskedastic Models, *Journal of Economic Dynamics and Control* **18** (1995), 931-955.

Evaluation of the locality influence on the electricity supplier selection

Sekničková Jana[1], Kuncová Martina[2]

Abstract. The electricity supplier selection in the Czech Republic started to be a complicated problem after the market transformation and deregulation in 2006. This situation caused the increasing number of suppliers and separation of the Czech Republic regions into three areas with different distributors. The selection of the appropriate supplier is influenced not only by the amount of the electricity consumption but also by the prices given by distributor and supplier. The aim of this paper is to select such a supplier that minimizes total costs of the electricity for given household and given tariff. The costs are dependent on the prices of low and high tariff and on the fixed monthly fee set by the supplier and by distributor. According to these conditions it is possible to use multi-criteria decision-making methods to find the best supplier. In this paper we use several methods such as WSA, TOPSIS or MAPPACC for the evaluation with respect to the locality.

Keywords: electricity market, suppliers, multi-criteria decision making, methods

JEL Classification: C44, C61, O13
AMS Classification: 90C29

1 Introduction

The growth of modern technologies and higher and higher usage of the electronic equipment to ease the work, to relax, to study, etc. causes the increasing demand for electricity. Although it is said that the new devices are energy-saving ones the electricity consumption rises. The transformation of the electricity market in the Czech Republic that started in 2002 (for households in 2006) has led to the increasing number of suppliers offering this commodity to households and companies. The retail electricity market consists not only of suppliers and customers but also of distributors, Energy Regulatory Office (ERU) and operator of the market (OTE) [7]. The Czech Republic is divided into three network parts operated by three distributors so the households can choose only the supplier. Each household has its tariff rate according to the supplier's conditions. The complete list of suppliers and their tariffs and prices is changing every year. The selection of the suppliers is dependent upon the price development but it can be limited by the contract conditions of the previous supplier. With respect to all these it is a hard task to find the best supplier. Various techniques and methods can be used to model the situation on the market [8]. In this article we use multi-criteria analyses for the products offered by all suppliers in 2014 for the given household. We also compare the suppliers and their prices in different regions.

2 Czech electricity market

The market economy theory is based on the supply and demand. Demand refers to quantity of product people are willing to buy at a given price. It is obvious for standard good that the higher price evocates the lower demanded quantity. Supply represents the amount of product suppliers are willing to sell when receiving a given price. This means that the higher price corresponds to the higher supplied quantity. We say the economy is at equilibrium when supply and demand are equal. The equilibrium price (and also the equilibrium quantity) is so reflection of supply and demand.

In the case of perfect competition in the market the economists assume that there are a number of different buyers and sellers in the marketplace and so the change of price responses to changes in supply and demand.

Also at the Czech electricity market suppliers (with their supply) and consumers (with their demand) meet. But it is difficult to speak about competitive market. There are a lot of consumers but with the number of suppliers (in economic notation) can be a problem. Electricity in the Czech Republic is taken via distributors and suppliers [7]. The problem is given by distributor – each region has only one distributor of electricity and in the energy market there are no substitutes. It means there is no competition between distributors in region. In a mar-

[1] University of Economics, Prague, Faculty of informatics and statistics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Repubic, jana.seknickova@vse.cz.
[2] University of Economics, Prague, Faculty of informatics and statistics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Repubic, kuncovam@vse.cz.

ket that has only one distributor of a good or service (in this case electricity), the distributor can control price. A consumer does not have choice of distributor, cannot maximize his or her total utility and has very little influence over the price of energy. In this case we talk about monopoly market structure. That price has to be regulated by a regulatory office (ERU).

The market of electricity within region can be called competitive as the number of suppliers is still increasing. The complete list of suppliers and their products and prices is not providing an easy survey. It is given by region (distributor) and also by the electricity take-off amount.

The price for the electricity is made by two parts. The first one is the controlled charge for services related to electricity transport from the generator to the final customer. This charge is annually given by Energy Regulatory Office [2]. It covers:

- monthly lease for the circuit breaker,
- price per megawatt hour (MWh) in high tariff (HT),
- price per megawatt hour in low tariff (LT),
- price per system services,
- price for the support of the renewable energy purchase,
- charges for the electricity market operator,
- value added tax (with electricity tax 28.30 CZK per 1 MWh).

The second part of the total price is given by the electricity supplier. It covers:

- fixed monthly fee for the selected product,
- price per megawatt hour (MWh) in high tariff (HT),
- price per megawatt hour in low tariff (LT),
- electricity ecological tax (34.24 CZK per 1 MWh including VAT).

The Czech Republic is divided into three regions operated by three distributors (Pražská energetická – PRE, E.ON, České energetické závody – ČEZ) so the households cannot choose the distributor (as was mentioned above) but only the supplier [8]. Each household has its tariff rate according to the supplier's conditions. Various tariff rates are offered to households but their selection is limited as there are limitations for the tariff usage. For this article the tariff rate D25d has been chosen as we use a bigger house as an example (it is assigned when the household uses electric water heater). In this case the household can use low and high tariff (LT and HT).

3 Data and methods

In this paper we compare electricity tariffs (alternatives) given by different suppliers from the consumer's point of view. The aim of this analysis is to look into dependency of locality influence on the electricity supplier selection.

This problem can be solved by several different approaches. The first one is probably the easiest. We can suppose the electricity consumption is non-stochastic, its amount is known and we can so calculate the total costs. The best tariff is the one with the lowest costs. This approach can be described by mono-criteria model and the results can be found via various web calculators (for example [2]). However electricity consumption is not given in real practice. We know average consumption and maybe standard deviation and so the better approach is using of stochastic models, such as simulation model. The third approach is the usage of multi-criteria (non-stochastic) model and comparing of alternatives with respect to several prices (criteria). As we have tried to use the second approach in [6], in this paper we discussed the first and the third denoted non-stochastic approach.

3.1 Data

According to the previous analysis [5] in this paper we use data for 57 tariffs (offered by 28 suppliers) in all three distribution areas. For illustration we selected 3 districts each from one area: Chrudim (for distributor ČEZ), Prague (for distributor PRE) and Pelhrimov (for distributor E.ON). All data come from web pages of the Energy Regulatory Office [2], they display prices for year 2014 and are calculated for the circuit breaker from 3x20A to 3x25A for each supplier for the tariff rate D25d. On the basis of the selected household in [5] the limits for the electricity consumption has been set for each month (the annual consumption is about 10 MWh, 45% energy in high tariff and 55% in low tariff). Each tariff was denoted by number 1 – 57 according to the position in list of alternatives.

Note that in area for ČEZ distributor data for all 57 tariffs are available on mentioned web pages. For PRE distributor data for two tariffs are unavailable (no. 53 – Pražská plynárenská Aku D and no. 54 – Pražská plynárenská Aku D sleva 3,5 %) and also for E.ON distributor data for one tariff (21 – Lumen Energy Fix 2014) are missing. These tariffs are not offered in given regions.

As we wrote in chapter 2, there are 11 items that influence the final price paid by consumer. However price per system services, price for the support of the renewable energy purchase, changes for the electricity market operator, electricity tax and electricity ecological tax are fixed for all tariffs, all suppliers and all distributor on condition that the total consumption is identical. Therefore for the analysis we use only six variables (six criteria): monthly lease for the circuit breaker, price of distributor per MWh in high tariff (HT), price of distributor per MWh in low tariff (LT), fixed monthly fee for the selected product, price of supplier per MWh in HT and price of supplier per MWh in LT. Note that values of the first three criteria are dependent only on distributor and so they are identical for all tariffs on condition that the distributor of energy is the same.

It is obvious a tariff that is dominated cannot be chosen as a winner and so for selection of the best tariff it is sufficient to analyze only non-dominated variants. As the second best alternative can be placed on the second rank in the case of final rank all alternatives have to be analyzed. Therefore the analyses were done for three groups of alternatives (tarrifs). The first one consists of all 57 alternatives, the second one consists of alternatives that are non-dominated with respect to at least one distributor (13 non-dominated alternatives in Table 1) and the third group is formed by alternatives that are non-dominated with respect to all three distributor (8 alternatives on the white background in Table 1).

no.	*marketing denotation*	*with respect to all distributors*
12	Centropol Energy eOptimum aku 8	dominated
13	Centropol Energy aktiv	non-dominated
17	Bohemia Energy Home Morava Basic	dominated
20	Lumen Energy home aku	dominated
29	ELIMON eProdukt	non-dominated
36	Europe Easy Energy duo	non-dominated
38	X Energie	dominated
39	Amper Market	non-dominated
40	Fosfa	non-dominated
45	Fonergy Premium	non-dominated
47	eYELLOW single	non-dominated
55	COOP Energy výzva2014 náš člen	dominated
57	CARBOUNION KOMODITY	non-dominated

Table 1 List of non-dominated alternatives

3.2 Methods

In the case we use only one criterion the objective function for analysis is the cost function. The total costs can be calculated for each distributor and each tariff according to formula (1):

$$COST_{ij} = (1 + VAT) \cdot \left[12 \cdot \left(mf_{ij} + mf_j\right) + c\left(0.45 \cdot \left(ph_{ij} + ph_j\right) + 0.55 \cdot \left(pl_{ij} + pl_j\right) + (os + t)\right)\right] \quad (1)$$

where
i … tariff, $i = 1, \ldots, 57$,
j … distributor, $j = 1, \ldots, 3$,
VAT … value added tax ($VAT = 0.21$),
mf … fix monthly fee,
c … yearly electricity consumption in MWh (in this study $c = 10$),
ph … price in high tariff per 1 MWh,
pl … price in low tariff per 1 MWh,
os … price for other services per 1 MWh,
t … electricity tax per 1 MWh ($t = 28.3$ CZK).

In the first step the value of this function is calculated for each tariff i and each distributor j. In the second step for each distributor j the tariff with the lowest value of cost function is chosen as the best tariff (BT_j),

$$BT_j = \arg\min_i COST_{ij}\,, j = 1, 2, 3. \quad (2)$$

In the third step with respect to increasing values of cost function all tariffs for given distributor can be sorted.

Multi-criteria evaluation of alternatives belongs to the category of discrete multi-criteria decision making models where all the alternatives ($a_1, a_2, \ldots, a_p$) and criteria ($f_1, f_2, \ldots, f_k$) are known. To solve this kind of model it is necessary to know the preferences of the decision maker. These preferences can be described by aspiration levels (or requirements), criteria order or by the weights of the criteria. We may find a lot of different methods

[3], [4], the three that we use are WSA, TOPSIS and MAPPACC. For each distributor we suppose three decision criteria (Price HT, Price LT and Fixed monthly fee) and 57 tariffs as alternatives (8 or 13 non-dominated alternatives respectively). As all used methods need cardinal information for criteria for the analyses we use weight vector **v** = (0.5, 0.225, 0.275) for fixed monthly fee, high tariff and low tariff (due to consistency with previous analyses [5] and [6]). Each suggested method will be applied on data for each distributor and then the results will be discussed.

4 Results and discussion

According to described methods we have started with analysis of the locality influence on the electricity supplier selection from the mono-criteria point of view. Afterwards we analyzed the same situation by multi-criteria methods.

4.1 Mono-criteria analysis

Firstly we have analyzed the problem from the mono-criteria point of view. For analysis we used all 57 alternatives. Table 2 displays the best five tariff with respect to total costs sorted by average rank. We can see that tariff no. 57 has the lowest total cost without reference to distributor. Also the second and the third rank is placed by tariff no. 29, and tariff no. 39 respectively, for each distributor. It is obvious that the best tariff from the formula (2) is the tariff no. 57 – Carbounion Komodity for all distributors and a model household will choose this tariff (with the lowest costs) with no respect to distributor and so with no respect to locality where it is placed. In this analysis we showed that the best tariff is independent on locality. It is obvious that the prices are different with respect to locality and household from Prague (distributor PRE) will have lower costs (30 703 CZK) than household from Pelhrimov (E.ON) with costs 30 782 CZK and from Chrudim (ČEZ) with 31 800 CZK. The change of locality influences total costs but not the best tariff.

distributor	E.ON		ČEZ		PRE		average rank
no.	costs	rank	costs	rank	costs	rank	
57	30 782,31	1	31 800,34	1	30 703,08	1	1,0
29	31 423,61	2	32 441,64	2	31 344,38	2	2,0
39	31 430,87	3	32 448,90	3	31 351,64	3	3,0
45	31 536,74	5	32 554,78	5	31 457,51	5	5,0
36	31 701,91	6	32 719,94	8	31 622,68	6	6,7

Table 2 Mono-criteria analysis

In this paper we have compared also all tariffs with respect to distributor (and also locality). We can show that 54 from analyzed 57 tariffs have the lowest total costs in the case of distributor PRE. Also 54 tariffs have lower costs in the case of distributor E.ON than in the case of distributor PRE. And finally there are 51 tariffs with the lowest costs for PRE and the highest costs for ČEZ in data file. The best tariff that does not satisfy this condition figures on seventh rank and it is the tariff no. 13 from table 1. For this tariff distributor E.ON offers cheaper alternative than PRE (31 493 CZK for E.ON versus 31 967 CZK for PRE).

It is also obvious that the total costs are directly dependent on distributor and also on locality where a model household is placed.

4.2 Multi-criteria analysis

The first analysis was done for only 8 tariffs that are non-dominated for all distributors. Results in table 3 show ranks with respect to WSA, TOPSIS and MAPPAC methods for each distributor. For analysis of dependency on locality we should compare only results obtained by the same method.

In the case of WSA using method the final rank is very similar for all distributors. The best tariff is no. 45, the second one no. 40 and the third one tariff no. 36. The change of locality does not influence selection of the best tariff.

Similarly in the case of TOPSIS the final rank is practically identical for all distributors and so the best tariff is independent on the change of locality. In the case of MAPPAC method the results are different but not significantly. The best tariff is also in this case no. 45, but the rank of tariff no. 36, and no. 40 respectively, differs. However the change of locality also in this case does not influence the best tariff. Note that tariffs no. 29 and 57 are in identical indifferent class.

method	WSA			TOPSIS			MAPPAC		
distributor	PRE	ČEZ	E.ON	PRE	ČEZ	E.ON	PRE	ČEZ	E.ON
1.	45	45	45	45	45	45	45	45	45
2.	40	40	40	40	40	40	36	40	40
3.	36	36	36	39	36	36	40	36	36
4.	39	39	39	36	39	39	39	39	39
5.	57	29	29	29	29	29	29	29	29
6.	29	57	57	57	57	57	57	57	57
7.	47	13	13	47	47	47	47	47	47
8.	13	47	47	13	13	13	13	13	13

Table 3 WSA, TOPSIS, MAPPAC for 8 non-dominated tariffs

Table 4 shows the results for the second analysis of 13 tariffs that are non-dominated for at least one distributor. The similarity with the results for 8 analyzed tariffs is obvious. The best tariff will be no. 45 with no respect to used method and also with no respect to distributor. Also the final rank is similar for all distributors although differences are more significant than in the table 3.

method	WSA			TOPSIS			MAPPAC		
distributor	PRE	ČEZ	E.ON	PRE	ČEZ	E.ON	PRE	ČEZ	E.ON
1.	45	45	45	45	45	45	45	45	45
2.	40	40	36	40	40	40	40	36	36
3.	36	36	40	12	36	36	39	39	39
4.	12	12	39	39	12	12	57	40	40
5.	39	39	12	36	39	39	36	57	29
6.	57	29	29	29	29	29	29	29	57
7.	29	57	57	57	57	57	12	12	12
8.	20	13	13	47	47	47	20	47	13
9.	55	20	55	20	20	13	45	13	55
10.	47	47	20	55	13	20	55	38	47
11.	17	55	38	17	55	55	17	20	20
12.	13	38	47	13	38	38	13	55	38
13.	38	17	17	38	17	17	38	17	17

Table 4 WSA, TOPSIS and MAPPAC for 13 tariffs

The third analysis applied WSA, TOPSIS and MAPPAC method on the whole data file of 57 tariffs. The results for the five best tariffs are displayed in table 5. Note that MAPPAC method in the case of PRE distributor denotes four indifferent tariffs – no. 29, 34, 35 and 57 – that could be placed in table 5 on the place (*). It is not surprising tariff no. 45 is again the best one without reference to method or distributor. The final rank differs more than in previous analysis however the selection of the best tariff is not influenced by distributor that is given by locality. This analysis also shows that non-dominated tariff has to be chosen as a winner.

method	WSA			TOPSIS			MAPPAC		
distributor	PRE	ČEZ	E.ON	PRE	ČEZ	E.ON	PRE	ČEZ	E.ON
1.	45	45	45	45	45	45	45	45	45
2.	36	36	36	12	40	40	36	36	36
3.	12	12	40	36	36	36	39	39	39
4.	39	40	8	40	12	12	12	34	34
5.	40	8	39	8	8	8	29*	35	35

Table 5 WSA, TOPSIS and MAPPAC for all 57 tariffs

5 Conclusion

Selection of the best electricity tariff can be a complicated task. The problem is caused by several factors. The first one is the dependency of total costs not only on electricity consumption but also on prices given by distributors and suppliers. The second one is the fact that electricity consumption is not deterministic and we are only able to estimate it for the next period. In this paper we have supposed fixed electricity consumption for a given household and analyzed the problem of the best tariff selection from the mono-criteria and multi-criteria point of view.

For analyses we have used all 57 electricity tariffs offered by suppliers in year 2014 in all three distributors' regions. The first analysis has used only one criterion – minimal total costs – and we have showed that the best tariff is no. 57 – Carbounion Komodity. This tariff generates the lowest total costs in region operated by PRE distributor as well as E.ON and ČEZ distributor. We have proven that the change of locality (in the area of given distributor and also outside this area) does not influence the best tariff.

The second analysis was free of fixed electricity consumption and three multi-criteria decision methods (WSA, TOPSIS and MAPPAC) were used for selection of the best tariff. Based on assumption non-dominance of the best alternative we have firstly analyzed data file with only eight non-dominated tariffs. This file was extended to 13 alternatives and finally we have analyzed all tariffs. In all cases no. 45 – Fonergy Premium as the best tariff was chosen. Note that this tariff was non-dominated for each distributor. Also here we have showed that the best tariff is independent on locality.

The prices of electricity differ between localities. If the household changes its place within one distributor's region the total costs stay unchanged and so the best tariff does not differ. If the household changes its place out of current distributor's region the total costs change with respect to different distributor's prices and also different supplier's prices in different distributor's regions. However in this study we have proven the best tariff is the same and tariff no. 45 – Fonergy Premium is the winner also in the case of locality change.

In this paper we have showed that in the Czech electricity market the selection of the best electricity tariff for a given household is independent on locality however the locality influences the total costs paid by household for electricity consumption. These costs are the lowest in the region of PRE distributor and are the highest in the region of ČEZ distributor in year 2014.

Acknowledgements

This work was supported by Grant No. P403/12/1387 of the Grant Agency of the Czech Republic and also by the grant IGA F4/19/2013 of the Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Banks, J.: *Handbook of Simulation*. USA, John Willey & Sons, 1998
[2] Energy Regulatory Office (ERU): *Price calculator* [online], available at: http://kalkulator.eru.cz/VstupniUdaje.aspx [cit. 2014-05-02]
[3] Fiala, P.: *Modely a metody rozhodování*. Praha: Oeconomica 2006.
[4] Figueira, J., Greco, S., Ehrgott M.: *Multiple Criteria Decision Analysis – State of the Art Surveys*. New York : Springer Science + Business Media Inc., 2005.
[5] Kuncová, M. and Rejmanová, V.: Hodnocení dodavatelů elektřiny pomocí matematických metod. In: *Hradec Economic Days 2014*. Hradec Králové : Gaudeamus, 2014, 123–129.
[6] Kuncová, M. and Sekničková, J.: Multi-criteria Analysis and Simulation Applied on the Czech Electricity Market. In: *Proceedings of Quantitative Methods in Economics Conference*, Virt, 2014, in print
[7] Roční zpráva o provozu ES ČR za rok 2012. *Energetický regulační úřad* [online]. [cit. 2013-11-15]. Dostupné z: http://www.eru.cz/user_data/files/statistika_elektro/rocni_zprava/2012/RZ_elektro_2012_v1.pdf.
[8] Ventosa, M., Baíllo, Á., Ramos, A. and Rivier, M.: Electricity Markets modeling trends, *Energy Policy* **33** (2005),.897–913.

Impact of cooperation of firms on both sides of the market on innovation

Ľubica Šimková
Institute of Public Policy and Economics, Faculty of Social and Economic Sciences, Comenius University in Bratislava
Mlynské luhy č. 4, 921 05 Bratislava
Slovak Republic
E-mail: lubica.simkova@fses.uniba.sk

Abstract. In our paper we analyze impact of two different modes of interaction between firms on product innovation of inputs. In both cases the decision on intensities of product innovation is made in the first period and interaction in setting prices of outputs follows in the infinite horizon. The first mode of interaction is characterized by Cournot duopoly in the market for final goods and Bertrand duopoly for market of inputs with differentiated product. Demand functions for inputs are derived from Cournot equilibrium in the market for final goods in which prices of inputs are parameters that producers of final goods cannot affect. In the second mode, all four firms cooperate with the aim to maximize the sum of profits. This applies to choosing intensities of product innovation in the first period as well as to subsequent setting of outputs in infinite horizon. The results indicate that cooperation between firms on both sides of the market has positive impact on intensity of product innovation of inputs. It also increases outputs of final goods and decreases their prices. Buyers of final goods benefit from this.

Keywords: product innovation, Bertrand duopoly, infinite horizon, maximization of sum of profits.

JEL Classification: C73, D43
AMS Classification: 91A20

1 Introduction

Product innovation of inputs leads to technological innovation of production of goods produced from these inputs (henceforth „final goods“). Intensities of such innovation depend on market structure in both input market and market for final goods. We address here the following question: Can cooperation between producers of inputs and producers of final goods enhance innovation of inputs in a way beneficial to buyers of final goods? We shed light on this issue by a simple example with two producers of inputs and two producers of final goods. The decision on intensities of product innovation is made in the first period and interaction in setting outputs follows in the infinite horizon. We compare Cournot equilibrium in the market for final goods and Bertrand equilibrium for market of inputs with differentiated product with cooperation between all firms aimed at maximizing the sum of profits. The results indicate that cooperation between firms on both sides of the market has positive impact on intensity of product innovation of inputs. Buyers of final goods benefit from this.

In research there are many authors focused on Cournot and Bertrand duopoly. To mention some of these authors, De Borger and Van Dender [1] focus on Bertrand duopoly. Dastidar [2] compared Cournot and Bertrand equilibrium in a homogenous product market. As well as Cournot and Bertrand duopoly, there exist many authors focused on innovations. Lambertini and Mantovani [3] studied process and product innovation by using a differential game. Cantner and Malerba [4] also examined innovation and industrial transformation.

2 Analyzed example

In our example we analyze impact of two modes of interaction between firms on product innovation of inputs. Both modes have in common that the decision connected with intensities of product innovation is made in the first period and interaction in setting prices for inputs and outputs of final goods follows in the infinite horizon. In the first mode, the non-cooperative decisions on intensities of input innovations in the first period are followed by the infinite repetition of the Bertrand equilibrium in the input market and Cournot equilibrium in the market for final goods. In the second mode, intensities of product innovations in the first period and outputs of both inputs and final goods in all periods are chosen in the way that maximizes the sum of the sums of discounted profits of all firms.

We work here with particular example. In this example we resume on work of Horniaček [5] but we analyze example with product innovation included. In the input market there exist two producers $(J=\{1,2\})$ and two buyers, who produce final goods from them $(I=\{3,4\})$. For each producer $j\in J$ cost function has the form $c_j(y_j)=10y_j+50$, where y_j is output of producer j. For each buyer $i\in I$ his total costs consists of expenditures on inputs and additional costs (e.g. labor cost, maintainance cost) expressed by function $c_i(x_i)=\sum_{j\in J}x_{ji}+20$, where x_{ji} is the quantity of input produced by firm j used by buyer i. Every $i\in I$ produces one type of good and its production function has the form $f_i(x_{1i},x_{2i})=\min\{(1+\Delta_1)x_{1i};(1+\Delta_2)x_{2i}\}$, where Δ_j is the intensity of product innovation of input produced by firm $j\in J$. The inverse demand functions for the buyers' outputs, $P_3:R_+^2\to R_+$ and $P_4:R_+^2\to R_+$, have the form $P_3(q_3,q_4)=\max\{42-2q_3+q_4,0\}$ and $P_4(q_3,q_4)=\max\{42+q_3-2q_4,0\}$, where q_3 and q_4, respectively, are outputs of final goods. The cost of product innovation by firm $j\in J$ with intensity Δ_j is $\frac{1}{5-\Delta_j}-\frac{1}{5}$. Thus, intensity of product innovation is from the interval $[0,5)$.

3 Infinite repetition of static equilibrium

In this case we analyze Nash equilibrium in intensities of input innovation between firms one and two in period one. We do so under the assumption that after making decision on intensities of output innovation the infinite repetition of Bertrand equilibrium in market for inputs follows. The latter is computed using demand functions obtained from Cournot equilibrium in the market for final goods.

In order to derive demand functions for inputs we first compute static Cournot equilibrium in market for final goods, depending on prices of inputs w_1,w_2 and on intensities of product innovations of inputs Δ_1,Δ_2. Firm 3 for given $q_4\geq 0$ maximizes the value of function

$$\Pi_3(q_3,q_4,w_1,w_2,\Delta_1,\Delta_2)=(42-2q_3+q_4)q_3-w_1\frac{q_3}{1+\Delta_1}-w_2\frac{q_3}{1+\Delta_2}-\frac{q_3}{1+\Delta_1}-\frac{q_3}{1+\Delta_2}$$

subject to $q_3\in\left[0,\frac{42+q_4}{2}\right]$. (1)

Firm 4 for given $q_3\geq 0$ maximizes the value of function

$$\Pi_4(q_3,q_4,w_1,w_2,\Delta_1,\Delta_2)=(42+q_3-2q_4)q_4-w_1\frac{q_4}{1+\Delta_1}-w_2\frac{q_4}{1+\Delta_2}-\frac{q_4}{1+\Delta_1}-\frac{q_4}{1+\Delta_2}$$

subject to $q_4\in\left[0,\frac{42+q_3}{2}\right]$. (2)

The Cournot equilibrium in this market is obtained from the solution of the system of equations generated by the first order conditions for (1) and (2). The solution is symmetric and is given by

$$q_3(w_1,w_2,\Delta_1,\Delta_2)=-\frac{1}{3\Delta_1+3\Delta_2+3\Delta_1\Delta_2+3}(w_1-41\Delta_2-41\Delta_1+w_2-42\Delta_1\Delta_2+\Delta_1w_2+\Delta_2w_1-40)$$

and

$$q_4(w_1,w_2,\Delta_1,\Delta_2)=-\frac{1}{3\Delta_1+3\Delta_2+3\Delta_1\Delta_2+3}(w_1-41\Delta_2-41\Delta_1+w_2-42\Delta_1\Delta_2+\Delta_1w_2+\Delta_2w_1-40)$$

Second order conditions for maximization problems (1) and (2) are satisfied. This gives following demand functions for inputs

$$y_1(w_1,w_2,\Delta_1,\Delta_2)=\frac{2}{3\Delta_1+3\Delta_2+3\Delta_1\Delta_2+3}x(-w_1+41\Delta_2+41\Delta_1-w_2+42\Delta_1\Delta_2-\Delta_1w_2-\Delta_2w_1+40)\frac{1}{1+\Delta_1}$$

and

$$y_2(w_1,w_2,\Delta_1,\Delta_2)=\frac{2}{3\Delta_1+3\Delta_2+3\Delta_1\Delta_2+3}(-w_1+41\Delta_2+41\Delta_1-w_2+42\Delta_1\Delta_2-\Delta_1w_2-\Delta_2w_1+40)\frac{1}{1+\Delta_1}$$

Using these demand functions we can now compute static Bertrand equilibrium in the market for inputs for given intensities of product innovations of inputs. Firm 1 for given $w_2\in\left(0,\frac{41\Delta_2+41\Delta_1+42\Delta_1\Delta_2+40}{1+\Delta_1}\right)$ solves the problem $\max Z_1(w_1,w_2,\Delta_1,\Delta_2)=w_1y_1(w_1,w_2,\Delta_1,\Delta_2)-10y_1(w_1,w_2,\Delta_1,\Delta_2)$ subject to

$$w_1\in\left(0,\frac{41\Delta_2+41\Delta_1-w_2+42\Delta_1\Delta_2-\Delta_1w_2+40}{1+\Delta_1}\right). \qquad (3)$$

Firm 2 for given $w_1\in\left(0,\frac{41\Delta_2+41\Delta_1+42\Delta_1\Delta_2+40}{1+\Delta_2}\right)$ solves the problem $\max Z_2(w_1,w_2,\Delta_1,\Delta_2)=w_2y_2(w_1,w_2,\Delta_1,\Delta_2)-10y_2(w_1,w_2,\Delta_1,\Delta_2)$ subject to

$$w_2\in\left(0,\frac{41\Delta_2+41\Delta_1-w_1+42\Delta_1\Delta_2-\Delta_2w_1+40}{1+\Delta_1}\right). \qquad (4)$$

The system of equations generated by first order conditions for problems (3) and (4) gives the following Bertrand equilibrium depending on intensities of product innovations of inputs. We get

$$w_1(\Delta_1,\Delta_2)=\frac{1}{3\Delta_2+3}(31\Delta_1+61\Delta_2+42\Delta_1\Delta_2+50)$$

and

$$w_2(\Delta_1,\Delta_2)=\frac{1}{3\Delta_1+3}(61\Delta_1+31\Delta_2+42\Delta_1\Delta_2+50).$$

Second order conditions for problems (3) and (4) are satisfied.

Now we can compute Nash equilibrium in intensities of product innovations in period one. We use discount factor $\delta=0.6$ for both firms. Firm 1 for given $\Delta_2\in[0,5)$ solves the problem

$$\max\Phi_1(\Delta_1\Delta_2)=\frac{0.6}{1-0.6}$$
$$\times(W_1(\Delta_1,\Delta_2)y_2(W_1(\Delta_1,\Delta_2),W_2(\Delta_1,\Delta_2),\Delta_1,\Delta_2)-10y_1(W_1(\Delta_1,\Delta_2),W_2(\Delta_1,\Delta_2),\Delta_1,\Delta_2))-\frac{1}{5-\Delta_1}+\frac{1}{5}$$

subject to $\Delta_1\in[0,5)$. Firm 2 for given $\Delta_1\in[0,5)$ solves the problem

$$\max\Phi_2(\Delta_1\Delta_2)=\frac{0.6}{1-0.6}$$
$$\times(w_2(\Delta_1,\Delta_2)y_2(w_1(\Delta_1,\Delta_2),w_2(\Delta_1,\Delta_2),\Delta_1,\Delta_2)-10y_2(w_1(\Delta_1,\Delta_2),w_2(\Delta_1,\Delta_2),\Delta_1,\Delta_2))-\frac{1}{5-\Delta_2}+\frac{1}{5}$$

subject to $\Delta_1\in[0,5)$. We focus on symmetric Nash equilibrium. We solve the equation obtained from the first order conditions for firm 1 after substituting Δ_1 for Δ_2. This equation has the unique solution in interval

$[0,5)$ which is $\Delta_1 = 4.4357$. For fixed $\Delta_2 = 4.4357$, the first order condition for firm 1 has again the unique solution in interval $[0,5)$ which gives the same value, that is 4.4357 . We have verified that this value of Δ_1 gives a higher value of the objective function of firm 1 $\Theta_1(4.4357, 4.4357) > \Theta_2(0, 4.4357)$. Thus, for $\Delta_2 = 4.4357$, the maximization problem of firm 1 has the unique solution $\Delta_1 = 4.4357$. Due to symmetry of the model, the analogous reasoning applies to firm 2.

Therefore, the Nash equilibrium in intensities of product innovation of inputs is $(\Delta_1, \Delta_2) = (4.43757, 4.4357)$.

This gives the following values of prices and outputs. Prices of inputs are $w_1(4.4357, 4.4357) = w_2(4.4357, 4.4357) = 78.766$.

Outputs of inputs are
$y_1(78.766, 78.766, 4.4357, 4.4357) = y_2(78.766, 78.766, 4.4357, 4.4357) = 1.5516$.

Outputs of final goods are
$q_3(78.766, 78.766, 4.4357, 4.4357) = q_4(78.766, 78.766, 4.4357, 4.4357) = 4.217$.

Prices of final goods are 37.783 and profit of a final producer in a single period is equal to 15.566 , in infinite horizon sum of his discounted profits equals to 23.349 .

4 Cooperation in intensities of innovations and the outputs

In this section we report the results of computation concerning the second mode of interaction between firms. We maximize the sum of discounted profits of all firms in infinite horizon, including cost of product innovation incurred in period one. In period one both producers of inputs make decisions on intensity of product innovation and then in period two, three (and so on) innovated inputs are produced and firms three and four use these inputs in production of final goods.

We have made computations for discount factor $\delta = 0.6$. In this case in each period $t \in N$, the sum of average discounted profits of all firms is maximized when the coalition of all four firms forms. This gives output of each final good equal to 19.094 production of each input equals to 6.6183 , each producer of final good uses amount 3.3092 of each input, the sum of single period profits of all firms equals 729.13 . Intensity of product innovation of each input is 4.7701 .

5 Conclusion

In our paper we have compared infinite repetition of static equilibrium in both markets, with cooperation between all firms aimed at maximizing the sum of profits. The results have shown that cooperation between firms on both sides of the market has positive impact on intensity of product innovation of inputs. These results mean that buyers of final goods benefit from such a cooperation. When we compare the results of mentioned two modes, we can state that cooperation is better for the final consumer. In the case of cooperation between firms output of each final good is 19.094 and in the second case it is equal only to 4.217 . Prices of final goods in a case with cooperation is 24 and in a case without cooperation is 37.783 . Intensity of product innovation of each input with cooperation between firms is 4.7701 and without cooperation is 4.4357 . We have verified that cooperation is beneficial to buyers of final goods. As we said above, we resume on work of Horniaček [5], whose subjects of interests are interconnected markets and maximized sum of profits without innovations. I focused on sustainability in my paper for the Conference 2014 of the Slovak Society for Operations Research [6], which is based on the work of Farrell and Maskin [7]. Despite the fact that we work here with particular example, it would be possible to generalize calculation procedure according to Horniaček [5] of course with innovation included. In our paper we decided to work with particular example because of the short range of the paper. In the future it is possible to work further on this theme and one of the possibilities is an identification of threshold, when positives of competition prevails.

Acknowledgements

This work was supported by grants VEGA 1/0181/12 and VEGA 1/0197/13.

References

[1] De Borger, B., and Van Dender, K.: Prices, capacities and service levels in a congestible Bertrand duopoly. *Journal of Urban Economics* **60** (2006), 264-283.

[2] Dastidar, K. G.: Comparing Cournot and Bertrand in a Homogenous product Market. *Journal of Economic Theory* **75** (1997), 205-2012.

[3] Lambertini, L., and Mantovani, A.: Process and Product innovation: a Differential Game Approach to Product Life Cycle. *International Journal of Economic Theory* **6** (2010), 227-252.

[4] Cantner, U., and Malerba, F.: *Innovation, industrial dynamics and structural transformation: Schumpeterian legacies*. Springer, Berlin; London, 2007.

[5] Horniaček, M.: *Cooperation and Efficiency in Markets. Lecture Notes in Economics and Mathematical Systems 649*. Springer, Berlin – Heidelberg, 2011.

[6] Šimková, Ľ.: An Example of Renegotiation-Proof Equilibrium with Product Innovation. In: *Quantitative Methods in Economics (Multiple Criteria Decision Making XVII)* (Reiff, M., and Gežík, P., eds.). The Slovak Society for Operations Research, Virt, 2014, 276-279.

[7] Farrell, J., and Maskin, E.: Renegotiation-Proof Equilibrium: Reply. *Journal of Economic Theory* **49** (1989), 376-378.

Stochastic Modelling of Age-specific Mortality Rates for Demographic Projections: Two Different Approaches

Ondřej Šimpach[1], Jitka Langhamrová[2]

Abstract. Nowadays it is not enough to construct the demographic projections based on the deterministic models. Stochastic models are more computationally intensive, but they are more expressive and may take account of many more factors, including e.g. random component. For the purposes of demographic projections it is necessary to know the expected future development of mortality, fertility and migration. In this paper we will show two different approaches to modelling age-specific mortality rates, and from the obtained models we prepare a projection of these rates to 2050. Data cover the period 1920–2012. The first type of model is an individual random walk with drift. Male and female population will be analysed for the range of 0–100+ years, and each completed age will be analysed as the individual time series with an annual frequency. The second model will be Lee-Carter, which is currently often used. It is based on Principal components method, which can capture and explain the main factors of mortality. Based on this models there will be constructed the forecasts of age-specific death rates for the period from 2013 to 2050. The obtained forecasts will be confronted and according to their provided results and their computational complexity there will be determined the recommendations about their suitability or their further improvement.

Keywords: population projection, age-specific mortality rates, random walk, Lee-Carter model.

JEL Classification: C32, C55, J11
AMS Classification: 60G25

1 Introduction

For a significant demographic projection it is necessary to know the possible future development of mortality. Mortality is an important component of population reproduction. Its level affects the length of life (Šimpach, Pechrová [19]). When we analyse the development of mortality, it is important to know that it is changeable during the human life (Gavrilov, Gavrilova [10]). The biggest changes come out at the highest ages (approximately 60 years and above), where the mortality has the different character in comparison with its character at lower ages (Keyfitz [14] or Thatcher, Kanistö, and Vaupel [20]). This is not only caused by small numbers of deaths, but also by small number of living at the highest ages. It is also necessary to realize that these data are affected by systematic and random errors. If we want to capture the most accurately mortality of oldest people it is good idea to make minor adjustments. This is mainly related to smoothing of mortality and possibility of its extrapolation until the highest ages. For smoothing we can use several existing models. Among the most famous are included Coale-Kisker model (see e.g. Boleslawski, Tabeau [3] or Gavrilov, Gavrilova [10]), Thatcher model or Kanisto model (see e.g. Thatcher, Kanisto, and Vaupel [20]), or the oldest one (but still used) is the Gompertz-Makeham function (see comparison study by Gavrilov, Gavrilova [10] and the application in Šimpach [18]). The disadvantage of these models is that *they cannot be used for projection of future mortality* and hence for the calculation of demographic projections. Demographic projections of possible future evolution of population are essential information channel, which is used for providing of important information about the potential evolution of mortality rates, birth rates, immigration and emigration, or other demographic statistics. Each projection is based on the assumptions, which could but might not occurr (Gardner, McKenzie [9]). Sophisticated stochastic demographic projections are based on the main components (Bell, Monsell [2] and Lee, Carter [15]), explaining the trend, which is included in the development of time series of age-specific demographic rates. The length of the time series has a major influence on results (see e.g. Coale, Kisker [7], or comparing the multiple results of populations from study by Booth, Tickle, and Smith [4]). In this article we focus on the evolution of the mortality in the Czech Republic. We use two approaches of stochastic modelling. The first one is the individual models of random walk with drift (see e.g. Bell [1]). We analyse the male and female population sepa-

[1] University of Economics in Prague, Faculty of Informatics and Statistics, Dept. of Demography and Dept. of Statistics and Probability, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, ondrej.simpach@vse.cz
[2] University of Economics in Prague, Faculty of Informatics and Statistics, Department of Demography, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, langhamj@vse.cz

rately for the range of 0–100+ years and the period from 1920 to 2012. We considered each completed age as the individual time series with an annual frequency of 92 observations. Based on 101 individual models for males and 101 for females respectively, we calculate the predictions up to the year 2050 (by Box, Jenkins [5] methodology). These forecasts are mainly influenced by trend, which is present in the evolution of time series. In addition to this approach we use the second one. Lee-Carter's model (Lee, Carter [15] or Lee, Tuljapurkar [16]) based on Principal components method can capture and explain the main factors of mortality. We estimate the parameters of that model for population of Czech males and females. We forecast the death rates up to the year 2050 using the main components of mortality. In the discussion we compare the obtained forecasts and according to their results and their computational complexity we determine the recommendations about their suitability or their further improvement. We determine pros and cons of these approaches in the conclusion.

2 Materials and Methods

For mortality analysis in the Czech Republic we use the data from the Czech Statistical Office (CZSO) about the number of deaths x-year-old $D_{x,t}^{M/F}$ (males and females separately), and the exposure to risk $E_{x,t}^{M/F}$, which is estimated as the midyear population x-year-old (males (M) and females (F) separately). We use the annual data from 1920 to 2012. The logarithms of age-specific death rates in population (e.g. Lee, Carter [15], Charpentier, Dutang [6] or Erbas et al. [8]) we calculate as

$$\ln\left(m_{x,t}^{M/F}\right) = \ln\left(\frac{D_{x,t}^{M/F}}{E_{x,t}^{M/F}}\right), \tag{1}$$

where $x = 0, 1, \ldots, \omega$, and $t = 1, 2, \ldots, T$. These empirical data can be seen in 3D perspective charts, (for R code see Charpentier, Dutang [6]) in Fig. 1, (where males are on the left side and females on the right side).

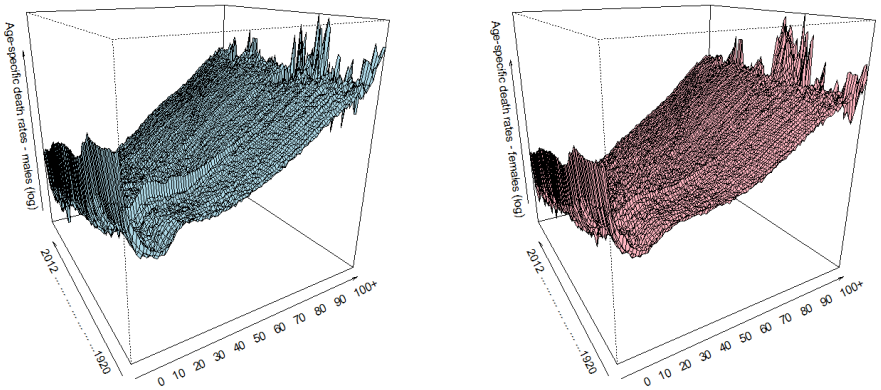


Figure 1 Age-specific death rates (in logarithms) of Czech males (left) and females (right) in 1920–2012. Source: CZSO, own construction and illustration in RStudio (R Development Core Team [17])

From the charts it is evident, that there were very unstable time series of death rates until 1948 (especially in the case of male population). It is well known that the instability of the time series reduces their predictive capability (Bell [1] or Gardner, McKenzie [9]). The history although has the lowest weight in the prediction model, but for the modelling of mortality, which is a long term process that has for each population its long-term trend, the history even with a little weight could be quite important (Booth, Tickle, and Smith [4]). In our paper we do not smooth the empirical data according to Gavrilov, Gavrilova [10] and rather use the random walk models with drift directly on raw data. (Lee-Carter model will be used on raw data as well). We denote Random walk model for male (M), female (F) population respectively as

$$\ln\left(m_{x,t}^{M/F}\right) = c_x^{M/F} + \ln\left(m_{x,t-1}^{M/F}\right) + \varepsilon_{x,t}^{M/F}, \tag{2}$$

where $c_x^{M/F}$ is constant and $\varepsilon_{x,t}^{M/F}$ is the error term with characteristics of white noise. This formula can be modified according to Box, Jenkins [5]. We get *ARIMA*(0,1,0) model with constant for male, female population respectively as

$$\ln\left(m_{x,t}^{M/F}\right) = \ln\left(m_{x,0}^{M/F}\right) + c_x^{M/F} \cdot t_x^{M/F} + \sum_{t=1}^{T} \varepsilon_t^{M/F} \,, \tag{3}$$

where $c_x^{M/F}$ $t_x^{M/F}$ is the deterministic trend. This trend is linear decrease of death rates in time. Our second used approach, the logarithms of age-specific death rates, can be decomposed (see Lee, Carter [15] or Lee, Tuljapurkar [16]) as

$$\ln\left(m_{x,t}^{M/F}\right) = a_x^{M/F} + b_x^{M/F} \cdot k_t^{M/F} + \varepsilon_{x,t}^{M/F} \,, \tag{4}$$

where $x = 0, 1, \ldots, \omega$, $t = 1, 2, \ldots, T.$, $a_x^{M/F}$ are the age-specific profiles independent of time, $b_x^{M/F}$ are the additional age-specific components determine how much each age group changes when k_t changes and finally $k_t^{M/F}$ are the time-varying parameters - the mortality indices. The estimation is based on Singular Value Decomposition (SVD) of matrix of age-specific death rates, presented e.g. by Bell, Monsell [2] and Lee, Carter [15]. For predicting the future age-specific death rates it is necessary to forecast the values of parameter $k_t^{M/F}$ only. This forecast is mostly calculated by *ARIMA*(*p*,*d*,*q*) models with or without constant (Box, Jenkins [5]). The values of the parameters a_x and b_x are independent of time and the prediction using the Lee-Carter model is therefore purely extrapolative (Lee, Tuljapurkar [16]).

3 Results and Discussion

Despite the fact that the random walk model with drift is relatively simple, it provides significant prediction, which involve the informations from the previous trend only. Neither additional expectations nor technological progress is implemented in results by random walk model (Coale, Kisker [7] or Hyndman et al. [11]). We estimated 101 male and 101 female individual models of random walk with drift (for particular series of age-specific death rates). Based on them we calculated the forecast of these rates up to the year 2050 in STATGRAPHICS Centurion XVI. From the output shown in the Fig. 2 it is clear that we can accept the general assumptions about the future declines of age-specific death rates, because the predicted time series for each year of life have a tendency to visibly decline in the future. In the case of the female population we can see only one irregularity at the age of 99 years. Due to the high variability of raw data (see Fig. 1) it is the predicted trend increasing at this age, but it is not so significant error.

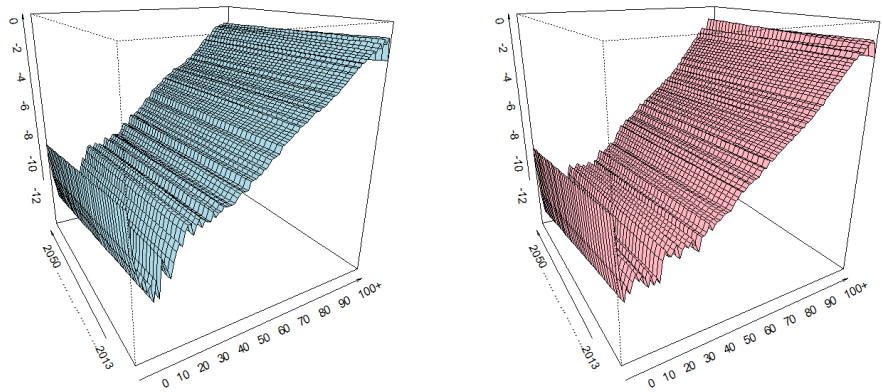


Figure 2 Forecast of age-specific death rates (in logarithms) for Czech males (left) and females (right) in period 2013–2050 by individual Random walk models with drift. Source: own calculation in STATGRAPHICS Centurion XVI, own construction and illustration in RStudio.

In the next step, using the SVD method implemented in the package "demography" (see Hyndman, [13]), which is developed for RStudio (R Development Core Team, [17]), we estimate the parameters a_x (age-specific profiles independent in time) and b_x (additional age-specific components determine how much each age group changes when k_t changes) for male and female Lee-Carter model. We can see these parameters in the Fig. 3, from which it is also clear the comparison between the different evolutions of parameters by sex. The age-specific profiles independent of time (a_x) are lower in the case of female model, because in general are the mortality rates of the female population lower in most age groups. The most significant difference between male and female mortality is in age group 18–32 years and at the oldest age groups. Higher mortality level of young males is caused by suicides, poisoning, dangerous behaviour, gambling, etc., (this is unfortunately a long-term trend).

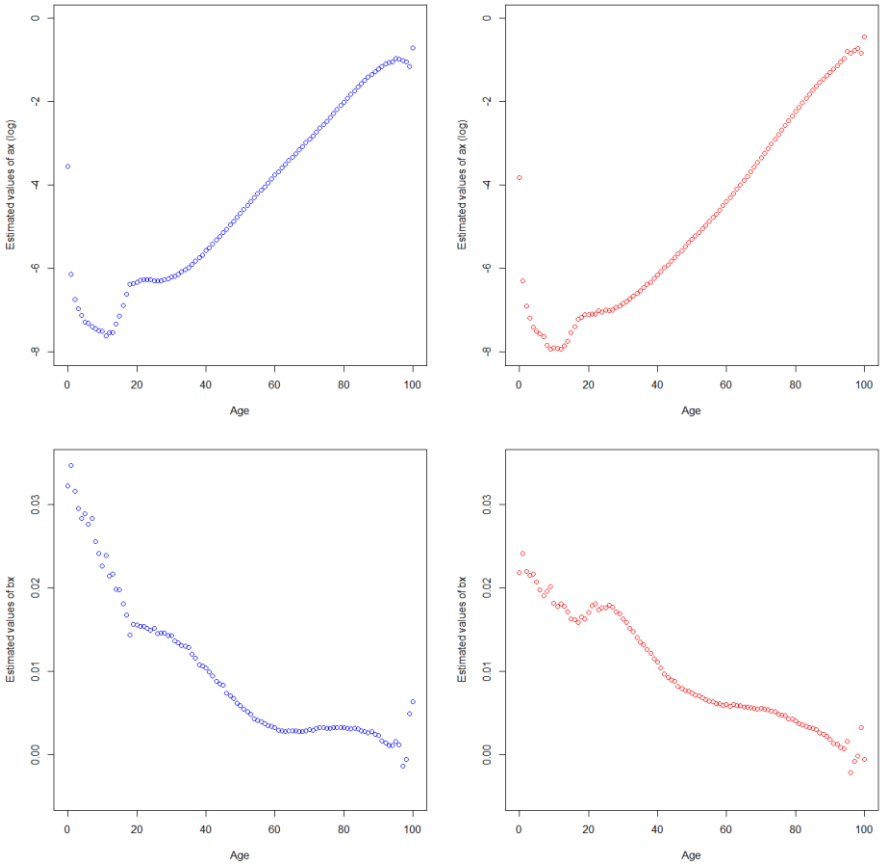


Figure 3 The estimates of age-specific profiles independent in time (parameter $\hat{a}_x$, top left for males and right for females) and the additional age-specific components determine how much each age group changes when k_t changes (parameter $\hat{b}_x$, bottom right for males and left for females). Source: own construction and illustration.

We also estimate the mortality indices k_t (the time-varying parameters) for the period 1920–2012 and these estimates are shown in the Fig. 4. To these estimates we calculate the predictions up to the year 2050 based on the *ARIMA* methodological approach, (Box, Jenkins [5]) and ran by "forecast" package in R (Hyndman et al. [11] and Hyndman, Shang [12]). Parameters of *ARIMA* models with drift are written in Tab. 1. From these predictions with 95% confidence intervals, (which can be seen in Fig. 4 too) it is clear that the model for female population provides slightly lower values of these estimates. Confidence intervals are wider in the model for male population (this is due to absence of AR parameter). Note that the theoretical basis for these predictions in sophisticated systems and automatic prediction software are prepared as well by Bell [1], Bell, Monsell [2] or Keyfitz [14].

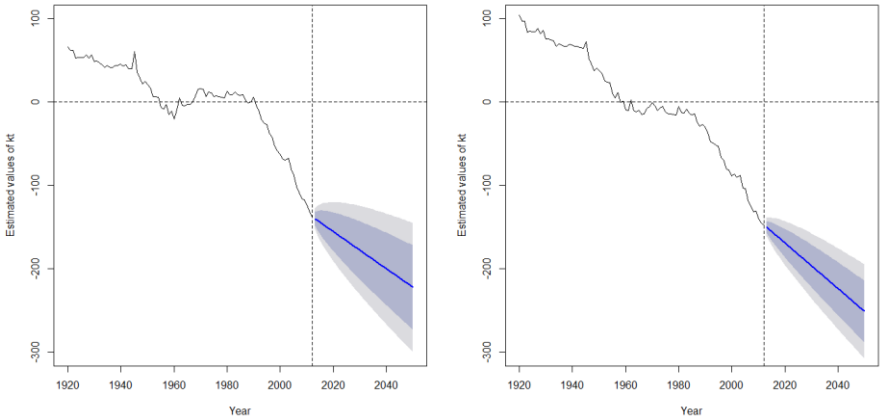


Figure 4 The estimates of the time-varying parameters $\hat{k}_t$ - the mortality indices with attached forecasts of these indices from 2013 to 2050 by ARIMA models. On the left side is the model for males, on the left side for females respectively. Source: own construction and illustration.

	AR(1)	s.e.	Drift	s.e.	AIC	BIC
Males ARIMA(0,1,0) with drift	x	x	-2.2181	0.6663	598.85	603.90
Females ARIMA(1,1,0) with drift	-0.2182	0.1020	-2.7256	0.4844	578.40	585.97

Table 1 Estimated parameters of ARIMA models for male and female $\hat{k}_t$. Source: own calculation.

Predicted values of the logarithms of age-specific death rates (from Fig. 2) by random walk models with drift can be combined with the empirical values of these rates (from Fig. 1). The result for male and female population is shown in the Fig. 5 (top). In order to compare these values all of the charts have the same scale. Based on the estimated parameters $\hat{a}_x$, $\hat{b}_x$ and $\hat{k}_t$ of two Lee-Carter's models we now fit and then estimate the future values of $\ln(m_{x,t})$ for males and females as

$$\ln\left(m_{x,t}^{M/F}\right) = \hat{a}_x^{M/F} + \hat{b}_x^{M/F} \cdot \hat{k}_t^{M/F}. \tag{5}$$

The result is shown in the Fig. 5 (bottom). Also in the case of Lee-Carter models it is clear that at the highest age groups break the assumption that the trend of the age-specific death rates decreases in the future.

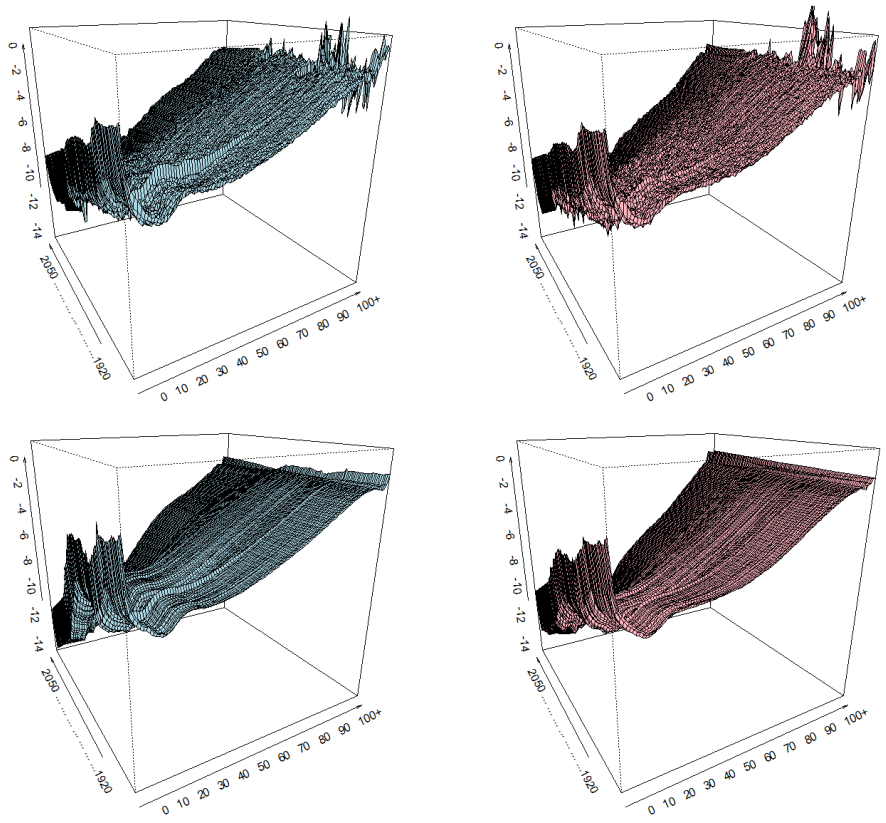


Figure 5 Age-specific death rates (empirical values in logarithms) of Czech males (top left) and females (top right) in 1920–2012 with attached forecasts of these rates from 2013 to 2050 by individual Random walk models with drift and age-specific death rates (fitted values in logarithms by Lee-Carter models) of Czech males (bottom left) and females (bottom right) in 1920–2012 with attached forecasts of these rates from 2013 to 2050 by Lee-Carter models. Source: own construction and illustration.

4 Conclusion

The aim of this paper was to show two different approaches to modelling age-specific death rates for male and female population. We estimated the models of random walk with drift for males and females. Then we forecasted and also estimated parameters of Lee-Carter model for both populations. Consequently we calculated the forecasts as well. In the case of the female population the results obtained by random walk models with drift and by the Lee-Carter model are almost comparable. Lee-Carter model just provides lower estimates of mortality rates at the lowest ages (which may be explained by the lower expected rate of future infant mortality). This result is fully consistent with study by Booth, Tickle, and Smith [4]. Differences depending on used model are

greater in the case of male population. Because we expect that the differences in male excess mortality will decrease in the future, it will be probably better to use the Lee-Carter model, which predicts a significant decline of logarithms of male age-specific death rates. Random walk models with drift are simpler, but in the case of male population we have to take into account also the additional information together with the past trend in time series. This additional information is significantly decreasing of the parameter b_x - the additional age-specific components determine how much each age group changes when k_t changes (see study by Lee, Tuljapurkar [16] and Hyndman et al. [11]). Results, which we obtained by Lee-Carter model, reflect better the expectations of some institutions. We can conclude that our predictions to a certain extent corresponds to the medium variant of the estimated future development of mortality in the Czech Republic, which periodically calculates the CZSO. In the conditions of the Czech Republic only the random walk models for the female population could be used.

Acknowledgements

This paper was supported by the project of Internal Grant Agency of VŠE Praha under the No. IGS F4/24/2013.

References

[1] Bell, W.R. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics*, 13(3): pp. 279-303.

[2] Bell, W.R., Monsell, B. (1991). Using principal components in time series modelling and forecasting of age-specific mortality rates. In: *Proceedings of the American Statistical Association, Social Statistics Section*, pp. 154-159.

[3] Boleslawski, L., Tabeau, E. (2001). Comparing Theoretical Age Patterns of Mortality Beyond the Age of 80. In: *Tabeau, E., van den Berg Jeths, A., and Heathcote, Ch. (eds.): Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*, pp. 127-155.

[4] Booth, H., Tickle, L., Smith, L. (2005). Evaluation of the variants of the Lee-Carter method of forecasting mortality: a multi-country comparison. *New Zealand Population Review*, 31(1): pp. 13-34.

[5] Box, G.E.P., Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco, Holden-Day, 537 p.

[6] Charpentier, A., Dutang, Ch. (2012). *L'Actuariat avec R*. [working paper]. Decembre 2012. Paternite-Partage a lindentique 3.0 France de Creative Commons, 215 p.

[7] Coale, Ansley J., Kisker, Ellen E. (1986). Mortality Crossovers: Reality or Bad Data? *Population Studies*. Vol. 40, pp. 389-401.

[8] Erbas, B., Ullah, S., Hyndman, R.J., Scollo, M., Abramson, M. (2012). Forecasts of COPD mortality in Australia: 2006-2025. *BMC Medical Research Methodology*, Vol. 2012, pp. 12-17.

[9] Gardner Jr. E.S., McKenzie, E. (1985). Forecasting Trends in Time Series. *Management Science*, 31(10): pp. 1237-1246.

[10] Gavrilov, L. A., Gavrilova, N. S. (2011). Mortality measurement at advanced ages: a study of social security administration death master file. *North American actuarial journal*, 15(3): pp. 432-447.

[11] Hyndman, R.J., Koehler, A.B., Snyder, R.D., Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3): pp. 439-454.

[12] Hyndman, R.J., Shang, Han Lin (2009). Forecasting functional time series. *Journal of the Korean Statistical Society*, 38(3): pp. 199-221 (With discussion).

[13] Hyndman, R.J. (2012). *demography: Forecasting mortality, fertility, migration and population data*. R package v. 1.16, http://robjhyndman.com/software/demography/

[14] Keyfitz, N. (1991). Experiments in the projection of mortality. *Canadian Studies in Population*, 18(2): pp. 1-17.

[15] Lee, R.D., Carter, L.R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, vol. 87, pp. 659-675.

[16] Lee R.D., Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: beyond high, medium, and low. *Journal of the American Statistical Association*, Vol. 89, pp. 1175-1189.

[17] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

[18] Šimpach, O. (2012). Faster convergence for estimates of parameters of Gompertz-Makeham function using available methods in solver MS Excel 2010. In: *Proceedings of 30th International Conference on Mathematical Methods in Economics, Part II*., Opava: Universita Opava, pp. 870-874.

[19] Šimpach, O., Pechrová, M. (2013). Assessing the impact of standard of living on the life expectancy at birth using Vector Autoregressive Model. In: *Proceedings of 31st International Conference on Mathematical Methods in Economics, PTS I and II*, Jihlava: College of Polytechnics Jihlava, pp. 921-926.

[20] Thatcher, R. A., Kanisto, V., and Vaupel, J. W. (1998). *The Force of Mortality at Ages 80 to 120*. Odense University Press.

Regional Input-Output Tables

Jaroslav Sixta[1], Jakub Fischer[2], Jaroslav Zbranek[3]

Abstract. Regional Input-Output Tables provide an extensive tool for economic analysis. These tables allow to deeply focus on selected regions and respect specific structural relationships rather than using national Input-Output tables. Regional models based on regional multipliers should provide better results that using national averages. Our estimates are based on multiregional model that provides appropriate detail of information with respect to necessary simplification. The process of compilation of regional Input-Output tables is usually demanding on data sources and costly and our papers describes own-developed procedures deriving regional tables and regional expenditure approach to gross domestic product. The key concept lies in the combination of existing input-output tables from national accounts with additional data from regional accounts and lots of auxiliary data sources. In the EU, information about regional value added is usually known but output and intermediate consumption have to be estimated, as well as inter-regional transactions. Finally, all the prepared regional input-output tables have to be balanced and they have to be fitted to national input-output tables.

Keywords: Input-Output, GDP, regional, accounts.

JEL classification: C67, R10, O11
AMS classification: 65C20

1 Introduction

Regional Input-Output tables (RIOTs) stay on the borderline between official statistics and research agenda. RIOTs are very often discussed in the literature [5], but there are only few practical examples. Construction of Regional Input-Output Tables requires a combination of both national and regional approach. Official statistics usually does not focus on RIOTs since the compilation is very demanding on data. Besides, due to the complexity of economic relationships, application of standard procedures used in national accounts is not possible. Lots of simplifications and assumptions have to be done to reach successful results. Direct and detailed data surveys of local units is impossible in recent times, these surveys are very expensive for both data providers (respondents) and compilers (statistical agencies).

Therefore RIOTs represent rather research issue than a standard statistical output. It is obvious that users of regional data would appreciate RIOTs but these tables should be used on the national level, as well. When we measure the impacts of different economic events, RIOTs should provide answers at least to two key questions. At first, what is the impact on the economy of the region and secondly, what are the relationships with other regions. In other words, when a large factory is being built in some region, different impact can be observed. For example, if there is not a local supplier of cement in the region, it has to be imported from another regions and overall effect of investments will be split into several regions.

2 Construction of Regional Input-Output Tables

Regular construction of Regional Input-Ouput Tables is very scarce even among the most developed countries. The most of existing RIOTs usually correspond to the group or block of countries rather

[1]University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, sixta@vse.cz
[2]University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, fischerj@vse.cz
[3]University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, jaroslav.zbranek@vse.cz

than regions of a single country. For example, pure regional tables are compiled by the U.S. Bureau of Economic Analysis [1] or Statistics Finland [11]. RIOTs had never been compiled for the Czech Republic and therefore we developed the methodology that combines both theoretical foundations and practical issues respecting Czech economic environment. Theoretical approach is well described in the literature, see [5] or [3] but there are only few practical compilations.

Unfortunately, European statistics is very connected with the definition of statistical unit in [4]. Since the main concept of a Local Kind-of-Activity Unit (LKAU) is not completely kept, the problems arises when RIOTs are going to be constructed, see [6]. Crucial issue lies in the decomposition of output and intermediate consumption of big enterprises. Formally, RIOTs can be described by Symmetric Input-Output Table in the same form as national one. In comparison with the Symmetric Input-Output Table (SIOT), key differences are found in foreign trade. Exports and Imports cover also exports and imports to and from other regions (including broad). There are two main concepts of RIOTs (see [5]) that determines their use:

1. Inter-Regional Model (IRM)
2. Multi-Regional Model (MRM)

Both concepts are possible to use and they differ in the recording of interregional transactions. IRM is based on the use of interregional trade and therefore it is very demanding on data. It is assumed that matrix of interregional-trade for each product and region is known. For r regions and i products, it is necessary to know rr number of matrices. On the contrary, Multiregional models assume existence of information of interregional trade of particular product. The destination of product (use) is neglected. Suppose matrices $\mathbf{Z}$ for intermediate consumption and $\mathbf{A}$ for technical coefficients. The difference between the models is reflected in technical coefficients, region specific technical coefficient for IRM can be expressed in formula (1)

$$a_{ij}^{rs} = \frac{z_{ij}^{rs}}{x_j^s}, \tag{1}$$

where a is a technical coefficient in region s for product i originated in region r used in production of product j, z is the intermediate consumption with the same indices and x stands for output of product j in region s. MRM is described by formula (2):

$$a_{ij}^{r} = \frac{z_{ij}^{r}}{x_j^r}, \tag{2}$$

where a is a technical coefficient in region r for product i used in production of product j, z is the intermediate consumption with the same indices and x stands for output of product j in region r.

Since pure Inter-Regional Models are very demanding for both data and computations, we used simplified MRM model. Our model is based on following assumptions:

1. Sum of all Regional Input-Output Tables correspond to the national SIOT.
2. Imports and exports between regions are estimated separately by auxiliary matrices and finally adjusted to fit production approach to GDP that is preferred.
3. Production approach to GDP and official construction of regional GDP by the Czech Statistical Office is completely respected.
4. Individual components of the expenditure approach to GDP are taken over from previous research done by our department, see [7].
5. RIOTs are compiled for all 14 regions of the Czech Republic (based on NUTS 3).

The above mentioned assumption means the national accounts published by the Czech Statistical Office are fully respected, see [10]. In the Czech Republic and in many other European countries, the

principle of LKAU[1] is not put into practice. The principal activity of the enterprise determines the activity and for the breakdown of big (multi-regional) company is usually done by top-down method or similar.[2] It means that interpretation of the results has to be adjusted. Suppose a company producing electricity. The company headquarter is located in the capital city but power plants are located in border regions. There are no transaction between the headquarter and power plants like selling the services of management, accounting, advertising etc. There are not even internal invoices within the company. Since LKAU principle is not kept, purchase of electricity in the capital induces production of electricity in border regions and production of services in the capital city. Services of the company produced in the capital are recorded under the code for electricity. It means that a customer buys both pure electricity and accompanying services but classified under the code for electricity.

Since we use Multi Regional Model, detailed trade transactions are omitted. RIOTs for 14 regions (NUTS 3 Kraje) are compiled with the respect to the national technology. In line [5] it is assumed that each product is produced by its specific way. Matrix $\mathbf{A}$ containing technical coefficients for total economy (from SIOT) is known and therefore elements z_{ij} of regional matrix Z^r are obtained as (3):

$$z_{ij}^r = a_{ij} x_j^r, \tag{3}$$

where x_j^r is output of product j in region r. Regional output is x^r is not published. We constructed output separately for market and non-market producers from annual national accounts. There were used similar approaches to regionalization of gross value added. In our model, regional Input-Output Tables formally correspond to SIOTs[3], see Figure 1 that describes general form of SIOT. There are lot of examples of both monetary and physical SIOTs (e.g. [12]) and the general form is also suitable for regional monetary Input-Output Tables.

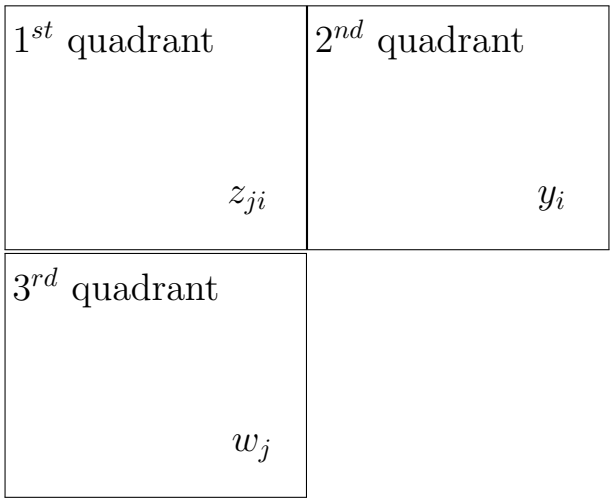


Figure 1 Structure of Symmetric Input-Output Table

The first quadrant describes intermediate consumption (z_{ji}) of product i in production of product j, the second quadrant describes final use (y_i) of product i and the third quadrant describes value added (w_j) in production of product j. For total output (x_j) of product j formula (4) is valid

$$x_j = \sum_i z_{i,j} + w_j. \tag{4}$$

[1] The principles of classification of units are given by [4] and [2].

[2] Pseudo bottom-up method is used, see [6]

[3] SIOTs are compiled according to Eurostat guidelines, see [2] and [8]. Product by product SIOTs have long tradition in the Czech Republic, see [8], [9].

3 Regional Output

Specifics of the regions can be described by regional output and value added. Regional output was derived by specific weights[4]. Regional value added and regional gross domestic product are usually published by the Czech Statistical Office. The quality and the detail of regional output substantially determine the quality of intermediate consumption and finally regional multipliers.

Estimated regional output is shown on the Figure 2. Substantial differences can be found between the regions. Capital city Praha is characterized by very balanced structure of output. The most important industries are connected with trade, manufacturing and energy, services and banking. On the contrary, Středočeský kraj and Moravskoslezský kraj have very different structure of output. In both regions, manufacturing plays a key role and other industries and substantially less important. It is clearly seen that manufacturing represent the most important industry for Czech economy. There are only two regions, Praha and Karlovarský kraj where manufacturing is not overshadowed by other industries. Very interesting is the situation of agriculture and forestry. From the perspective of value added, the importance is very low both at national level and regional level. In some regions (Praha, Karlovarský kraj, Liberecký kraj), the agriculture and forestry is nearly negligible.

Regional approach to output instead of value added allows us to determine regional specifics. Existence of regional output is a necessary condition for construction of regional multipliers that represent the crucial benefit of RIOTs. Since the measurement of output in national accounts is very complicated, regionalisation is very demanding. It is also connected with national approach to KAU. It is obvious that better definition of KAU leads to higher quality of RIOTs.

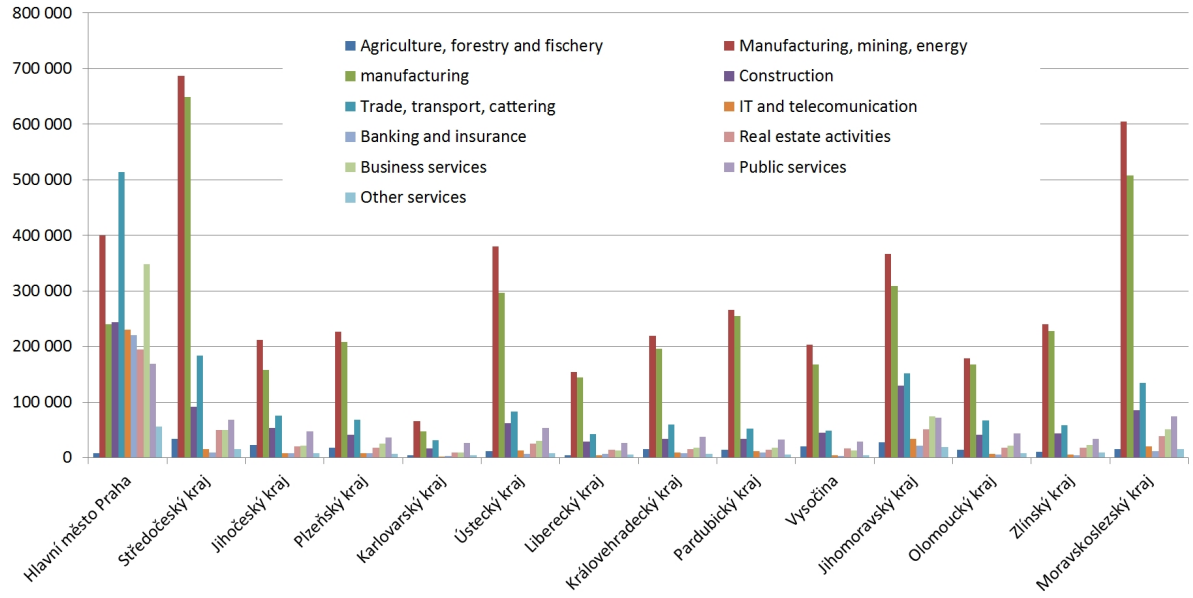


Figure 2 Output Multipliers, 2011, CZK mil.

4 Regional Multipliers

Regional Input-Output Tables were constructed for 14 regions of the Czech Republic according to the briefly described procedure above. The crucial issue is whether the regional multipliers differ between the regions. Since the output is specifically distributed across the regions, corresponding intermediates should be as well.

Our RIOTs have not been fully completed up to now. Currently we finished the first and partially

[4]The issue of construction of regional output goes far beyond the possibilities of this paper. It was a long process successfully finished with the help of the Czech Statistical Office, namely Mr. Kahoun who provided regional weights for industry totals.

the second quadrant of RIOT[5] The third quadrant covering components of value added is not completely finished. The problem lies in the consistency with officially published figures for gross value added. Construction gross value added as a difference of output and intermediate consumption estimated according to formula (5) leads to different figures. Therefore the first estimate of output multipliers has to be regarded as provisional. When RIOTs are ready, regional gross value added in each RIOT will correspond to published figures.

Both national and regional output multipliers are derived according to formula (3) and they express the impact of the increased final demand (Δy_i) for the additional product by 1.

$$\mathbf{m} = \mathbf{i}^{\mathbf{T}}\mathbf{L}, \tag{5}$$

where $\mathbf{m}$ is a vector of output multipliers,$\mathbf{i}^{\mathbf{T}}$ represent vector with unit elements (1,1,1,) and L is Leontief inverse, $\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1}$.

Output multipliers should be estimated for domestic and imported intermediates separately. Since we had only total RIOTs, the first estimates are based on provisional split of total use between domestic and imported intermediates. Computed multipliers for all 14 regions are shown in Figure 3. The interpretation of multipliers is following. For example, if final demand is increased by 1 for all products in all regions, it induces the increase of output by 1.876 in the Czech Republic. There are clearly seen different groups of rich and poor regions. Top value is observed for Jihomoravsk kraj where the increase of demand by 1 CZK creates the output of 2.05. On the contrary, in the region Vysočina additional demand for 1 CZK creates less than 1.7 CZK of output. It is clear that the region of Praha (the richest region, see. [6]) is very close to the average (1.876).

The importance of output multiplier lies in the possibility to react for regions specific. The increased demand for a particular product in some region leads only to the increase of import instead of stimulation of regional economy. For example, the increased demand for agriculture products in the capital city usually stimulates imports from other regions since these products are rarely produced in the capital. A similar situation is in the case of construction works where workers usually travel around the country for established construction places. Besides, induced production can raise imports since some production is based on imported intermediates. Increased demand for furniture will cause increase production of timber as well as food production and catering needs agriculture inputs etc.

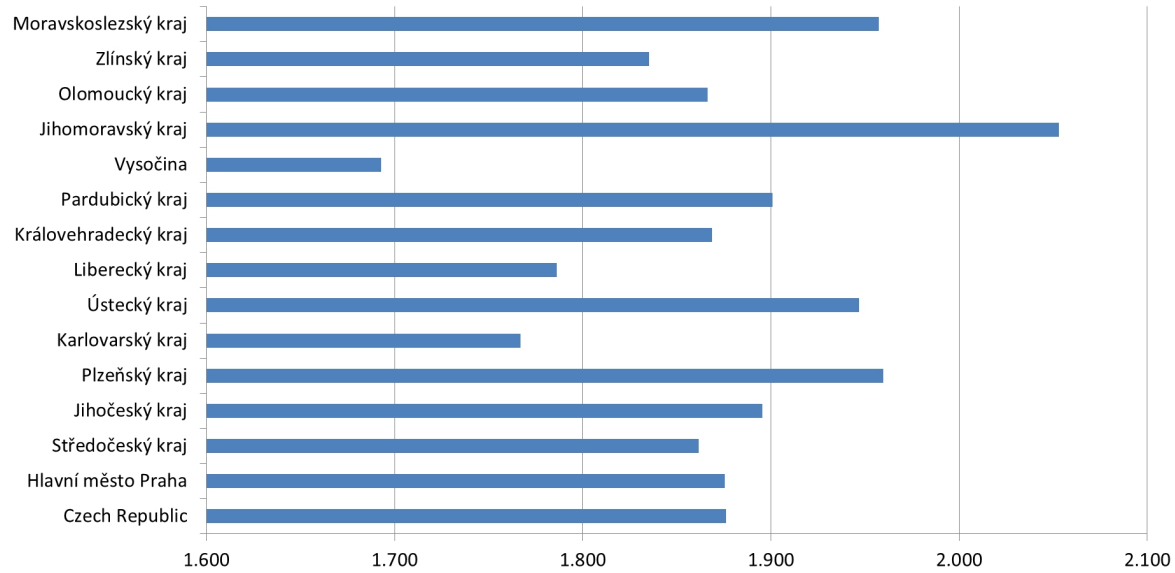


Figure 3 Output Multipliers, 2011.

[5]Up to now, we published only rough estimates in [10]. We intend to finish whole project by the end of 2014.

5 Conclusion

The most important benefits of Regional Input-Output Tables consist in the provision of the picture of regional relationships. There are usually very few practical examples of regional tables even though theoretical approach is very well described. Regional accounts are regularly published by official statistical agencies, e.g. the Czech Statistical Office in the Czech Republic. Official regional accounts cover only gross value added, employment, disposable income and gross capital formation. They do not cover output, intermediate consumption and both complete expenditure and income approach to GDP, see [6].

Construction of regional tables is still rather research agenda than official statistical output even it is not completely new issue. This will not be changed in the near future and therefore users have to rely on research projects. Our RIOTs are based on the combination of officially published data and own estimates. These tables are compiled for 14 regions in dimension of 82×82 products for 2011. Up to now, we finished provisional estimates of complete tables with simple allocation of import. Therefore we are able to compute output multipliers for all regions. There are significant differences between regions. The most sensitive regions for the change of final demand are Jihomoravský kraj and Plzeňský kraj. On the contrary, the Vysočina is the least sensitive region. Capital city of Prague is in the average.

The key issue lies in the use of detailed regional data. Regional analysis should serve to the creation and assessment of regional policy. Especially in recent times, regional unemployment represents significant problem that will need a complex approach in the whole EU. Tax incentives, government purchases or other means of support can be planned also with the help of Input-Output analysis. Regional Input-Output Tables provide the most detailed regional data about relationships that can be used for detailed analysis aimed at the measurement of impacts of regional support.

Acknowledgements

This paper is prepared under the support of the project Regional estimates of gross domestic product based on the expenditure approach of the Czech Science Foundation, project No 13-15771S.

References

[1] Bureau of Economic Analyses. [online]. URL: https://www.bea.gov/regional/rims/rimsii/

[2] European Union: *Manual on regional accounts methods*. Luxembourg, 2013.

[3] European Union: *Eurostat Manual of Supply, Use and Input-Output Tables*. [online], 2008. URL: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-013-EN.pdf

[4] Eurostat: *European System of Accounts - ESA 1995*. Luxembourg, Official Publications of the European Communities, 1996.

[5] Miller, R. E. and Blair, P. D.: *Input-Output Analysis: Foundations and Extensions*. Cambridge University Press, New York, 2009. ISBN-13 978-0-511-65103-8.

[6] Kahoun, J. and Sixta J.: Regional GDP Compilation: Production, Income and Expenditure Approach, *Statistika* **4** (2013), 24–36.

[7] Kramulová, J. and Musil, P.: Experimentální odhad složek výdajové metody regionálního HDP v ČR, *Politická ekonomie* **6** (2013), 814–833.

[8] Sixta J.: Development of Input-Output Tables in the Czech Republic, *Statistika* **2** (2013), 4–14.

[9] Sixta J. and Vltavská K. and Fischer, J.: The Development of Gross Domestic Product in the Czech Republic and Slovakia between 1970 and 1989, *Ekonomický časopis* (2013), 549–562.

[10] Sixta J. and Vltavská K. and Fischer, J.: Compilation of Regional Input-Output Tables. In: *Applications of Mathematics and Statistics in Economy (AMSE 2013)*, Banská Bystrica, 2013, s. 23.

[11] Statistics Finland. Online. [online]. URL: https://www.stat.fi/til/apt/index_en.html

[12] Zbranek J. and Sixta J.: Possibilities of Time Input-Output Tables. In: *Mathematical Methods in Economics 2013*, Jihlava, 2013, 1046–1051.

Comparative sensitivity analysis of CGE model results obtained by different model formulations

Vaclav Skoluda[1],

Abstract. CGE models belong to a broad family of general equilibrium models. These are large static models used for empirical analysis of macroeconomic shocks. The original model VŠEMOR2 and adapted model STDCGE represent tools for obtaining results. They focus mainly on the comparative analysis of selected countries and the sensitivity analysis of results. The proposed paper compares obtained results by these two models for different formulations and solving procedures. Models are formulated and solved as model of nonlinear programing (NLP), mixed complementarity problem (MCP) and constrained nonlinear system (CNS). These three approaches to solving of computable general equilibrium models are discussed and results are compared for more than 20 European countries. The proposed paper is based on author´s dissertation thesis and stands next to his other papers.

Keywords: CGE models, general equilibrium, comparative analysis, sensitivity analysis, nonlinear programing, mixed complementarity problem, constrained nonlinear system.

JEL Classification: C68, D58
AMS Classification: 91B5

1 Introduction

The aim of the presented article is to compare the solutions of computable general equilibrium (CGE) models on a conceptual level. On two different models compares the consequences of using different solving methods. For the analysis are used two models: VŠEMOR2 model, which was introduced in detail for example in [13] or [17] and model STDCGE [9], which belongs to the example models in GAMS environment. Different ways of solving the models consisted in formulation of the models as Nonlinear programming (NLP), Mixed complementarity problem (MCP) and Constrained nonlinear system (CNS). The solution was carried out through a number of solvers, CONOP, MINOS, SNOPT and PATH.

The results are analyzed on the example of 23 European countries selected by data availability. Main source were Input-Output tables published by Eurostat. All datasets in the form of Social accounting matrices (SAM) are created in relation to base year 2005, which was the last with available data for as many countries. More about creating SAM is in general [9] and in particular [15] and [16]. In CGE models we follow the change between two equilibrium states, before and after the shock to the economy, this change is expressed as a percentage. Shock in our case, represents an increase in the tax burden, expressed as the relative increase in government tax revenues by three percent. The analysis was composed with the *ceteris paribus* rule in mind, in order to compare the results with each other. The input data for the particular countries for both models are the same and the same is also the formulation of introduced shock, so the models are different in inner structure and thus its results.

2 Model VŠEMOR2

Presented model VŠEMOR2 is the second evolutionary level of the original model. Changes have been made particularly with regard to the input data and the ability to compare results with the example model STDCGE. Author's intent was to create such a model, which allows monitoring the behavior of the economy through two CGE models with the same input data and the same scenario for as many countries as possible. The model consists of 29 blocks of equations, which are aggregated according to the number of sectors in the SAM converted into tens or hundreds of individual equations. The equations are based on microeconomic theory and describe the demand of firms, households, government and virtual investment agent, abroad, balance of payments, express zero profits, income of households, government and investment and represent balance on the markets. The model itself and its code are based inter alia on the works [9], [12], [22] and the results were presented in [13], [15] and [17].

[1] University of economics, Prague, Department of econometrics, W. Churchill sq. 4, Prague, vaclav@skoluda.sk.

2.1 Solving model as NLP

The first presented way to deal with the CGE model is probably the most common approach through the formulation of the model as a nonlinear programming. In this case, the model equations become the constraints and the objective function is either artificial or appropriately selected one of the equations. The model is solved by three different nonlinear solvers in GAMS environment, namely CONOP, MINOS and SNOPT. Each is based on a slightly different principle of seeking solutions to systems of nonlinear equations, but all are without problems able to provide a solution. The table compares the effects of the scenario between three different solving methods for all available countries. The data in the rows marked with an asterisk are mentioned for completeness. Solving the model for the Netherlands led to a division by zero so all the results is irrelevant. For other countries asterisk means that solver MINOS identified solution as infeasible.

Single look at the data in the table indicates a relatively high sensitivity of the model to the solver used. Country by country, the results differ from each other. For example, in Finland the difference between the impacts on the economy after the introduction of shock in the form of tax increase by three percent for different solvers is minimal. On the other hand, is for example the Czech Republic in which the differences between the solutions obtained by the solver SNOPT and CONOP not negligible. As an aid can in this case serve the solution of the model as mixed complementarity problem.

According to the results in the table it looks as if each solver in GAMS had its own view on the economy, and its response to the introduction of shock. In that what is important they conform and the other is a matter of approach. In most cases, with a deviation of about tenths of percent, all solvers conform that tax increases will increase the total government consumption. This is also evident from the economic point of view. Impacts on other areas look as if solvers have held different economic theories. This is mainly about the resources to the increase of government consumption. In this case is CONOPT favoring consumption of investment, so the source is savings in economy. The other two solvers prefer dragging the government consumption by household consumption and foreign trade. The results are the more interesting they can be both economically interpreted and is the preference of the analyst for which interpretation he decides. From theoretical point of view, results are equal, both are locally optimal and they differ only in solving way to this local optimum.

The situation mentioned is not very pleasant for the analyst, who prefers clear solutions. Three different solving methods give three different results, which are mathematically equally good. The analyst is thus faced partly philosophical question of whether the problem is in the model or in the solution process. The model can of course be adjusted to give for all algorithms the same solutions, but only at the price of complexity in details of the model. It is preferable to former recognition of the fact that it is a non-linear optimization with non-convex set of solutions so that this situation can be expected in complex models, such as CGE models.

2.2 Solving model as MCP

An alternative way of solving the computable general equilibrium model is its formulation in the form of a mixed complementarity problem. A non-linear model is a square set of equations, where the number of equations and the number of variables is the same. Furthermore, it is necessary to identify the complementary pairs of equations and variables. As with the formulation of the CNS it's a formulation without artificial objective function. As an example of model solved in the form of MCP can be used [20].

Practical change from NLP to MCP is very simple. In essence it is enough to simply remove the objective function, which is usually in computable general equilibrium models artificial, so at this stage it do not lose any information from the model. In addition to the objective function, it has to be removed also its corresponding parameters and variables and modified the code to solve the model through the MCP. Solving the actual model takes place in GAMS environment and the analyst is there to interpret the results. Here we come to probably the biggest disadvantage of solving through MCP, which is low feedback from the solver in case of problems with solving the model. In solving in the form of NLP, the solver provides much additional information related to the solution to the problem. If the problems were not difficult to locate its source, it's easy to solve them. Formulation and solution by MCP puts higher demands on the analyst and in particular to his mathematical knowledge to solving nonlinear problems. This may be one reason for lower popularity of this solution method among analysts. The great advantage of such a formulation of the problem is the high speed of calculation, which is reflected in the tasks of larger dimensions.

Look into the table with results indicates minimal variation in the case of formulation of the model as a mixed complementarity problem. An asterisk in solutions appears only in the case of Hungary and Turkey, where solver reported an error so the solution is not in this case relevant and in the table is only for completeness. The actual results do not show major differences from other solving methods. Theoretical assumption of increasing total government consumption expenditure as a result of the tax increase was confirmed in all the

countries. In most cases, the results obtained through NLP solver SNOPT and after the reformulation of the model as MCP are closest. France or Austria could serve as an example. However, in several cases the results differ from the other three obtained by the solution as NLP. An example would be Portugal or Romania, where the change in government consumption is accompanied by increased household consumption. Solving through the formulation of the model as MCP again brings analyst to the crossroads just as was the case through NLP solvers.

For the completeness, all the results for the MCP were obtained through the PATH solver. It is possible to use alternative solvers, MILES and NLPEC. The first one has problems with multiple solution of model in scenario analysis and the other reformulate the model as NLP and further use of any of the other solver, the default is CONOP and in this case, then the results are identical.

2.3 Solving model as CNS

The second alternative to solve the CGE models other than by NLP is a reformulation of the model as Constrained Nonlinear System (CNS). Models are usually formulated in a square shape, or without major problems can be in this shape completed. The advantage of this is no need for artificial objective function and its verification in the system of nonlinear equations. The disadvantage, on the other hand, is weaker algorithm implementation, although only a few solvers can solve such formulated tasks. This approach is also not suitable for the analysis of large models, because in this case, may occur that some of equations in system will come near their linear combinations, resulting in the model will not be squared and cannot be solved. This can be easily fixed in the preparatory phase by detailed analysis of the model equations. The great advantage of this solution method is its calculation speed, which is especially evident by large models, such as [5].

In practice is solution by CNS very similar to solution through MCP. From the finished model NLP, just remove the objective function and instruct the program to solve it as CNS. All results are obtained by solver PATH, whereas the second solver, which was CONOP, did not provide a solution in many cases. Comments on the results are again similar to that of the MCP. Theoretical assumption about the increase in total consumption of government due tax increase is confirmed. Other values are more or less identical to the solution through the MCP, which can partly be given by using the same solver. The big differences are only in Denmark, France, Portugal and Sweden.

3 Model STDCGE

Second model is STDCGE, through which we observe the changes in the results due to different resolution procedure. It is a multi-sector open economy model made by N. Hosoe and described in detail in [9], on which is also based this analysis. Featured model had to be slightly modified to be able to use the same input data and the same scenario as in the model VŠEMOR2. In this section we will follow the same three approaches to model solving thru NLP, MCP and CNS.

3.1 Solving model as NLP

Insight into the table shows two major facts. The first is the low ability of solvers to achieve a solution, while the worst among them is MINOS. The second fact is that when more solvers provide results they in most cases match. All solvers provided solutions only in the case of Denmark and France and these solutions were the same. On the other hand, there are countries like Greece, Ireland and the United Kingdom, where any of the solvers failed to provide a solution without additional bug report. A similar situation occurred in the case of Lithuania, where the solver SNOPT felled in loop. Data for the Netherlands are omitted, since none of the solvers were able to get through division by zero. The results for countries in which the solvers were able to reach the equilibrium state comply with the economic assumptions. The only exception is Turkey.

This consistency contrasts with the results of the model VŠEMOR2 where the results among solvers were more divided. On the other hand by the model were VŠEMOR2 obtained results for more countries. This may be due to several factors, or a combination thereof. The models consist of different equations sets with different description of complexity of processes in the economy. Models provide to solvers starting values for finding baseline equilibrium and these values are in its area. Given nonconvexity of solutions set can this lead to restrict the model to find local optima and on the other hand, the excessive distance of starting values can guide the solver in another direction. Consistency with economic theory occurred in the same sense as in previous cases that tax increase resulted in increased total government consumption expenditure. In most cases, this change is accompanied by a slight decrease as in the total household expenditure on consumption so in foreign trade.

3.2 Solving model as MCP

Pros and cons of this approach in this case are the same as in the previous model. Problems with the solution occurred in several cases, which are marked with an asterisk in the table. PATH solver is not very communicative and limited to the simple message "*other error*". More detailed examination of errors usually lead to the disclosure of the problem in the form of a square root of a negative number, to which the solver reached during the iterative improvement and the results after removal of the problem occurred with another slow improvement of the results, thereby solver reached the limit value and thus did not improve the current solution that was not feasible.

The actual solutions are in most cases identical with the solutions obtained by one of the other ways forward. In several cases are more or less different, as an example can serve Belgium, Romania and Slovenia. For Great Britain other solvers failed and the only solution is through the MCP. More detailed view of the results shows however no change in foreign trade and production, which does not match the expectations and so these results do not pass the economic verification. Similar problems have for this solution process the results for Belgium.

3.3 Solving model as CNS

Calculation through the CNS showed for the countries least number of error messages of the three methods observed. For the solution was again used PATH solver for the same reasons as in the model VŠEMOR2. The results for each did not much differ between solution methods. Cases marked with an asterisk indicates a problem in the calculation and given the same solver as in in the previous section for the MCP, the reasons for failure are similar. Worth a look are also the results for Belgium, which are identical for NLP / SNOPT and CNS, but different from the MCP, while CNS and MCP uses the same solver. A similar situation applies in the case of Greece, where the solution through NLP completely failed. Results provide only alternative methods of calculation, with minor variations between them. For Ireland provided a solution by CNS as the only results without error message. Closer look again indicates problem with the economic verification of results.

		VŠEMOR2					STDCGE				
		% change in total consumption of Households	% change in total consumption of Government	% change in total consumption of Investment	% change in total Import	% change in total Export	% change in total consumption of Households	% change in total consumption of Government	% change in total consumption of Investment	% change in total Import	% change in total Export
AT	conopt	-0,18	0,78	-6,77	-2,29	0,24	-0,45	2,11	-0,86	-0,32	-0,31
	minos	-0,12	0,78	-0,59	-0,15	-0,15	2828968 *	-78,13 *	-6648098 *	-103,04 *	-100 *
	snopt	-0,11	0,78	0,16	0,12	-0,19	-0,45	2,11	-0,86	-0,32	-0,31
	CNS/PATH	-0,14	0,78	-1,96	-0,63	-0,07	-0,45	2,11	-0,86	-0,32	-0,31
	MCP/PATH	-0,14	0,78	-2,04	-0,66	-0,06	-0,45	2,11	-0,86	-0,32	-0,31
BE	conopt	-0,23	0,75	-2,82	-0,72	-0,25	-0,54 *	2,13 *	-1,2 *	-0,48 *	-0,45 *
	minos	-0,03	0,74	16,63	3,25	-1,16	-0,54 *	2,13 *	-1,21 *	-0,47 *	-0,44 *
	snopt	-0,19	0,74	3,35	0,5	-0,57	-0,54	2,13	-1,2	-0,48	-0,45
	CNS/PATH	-0,23	0,75	-4,25	-1	-0,17	-0,54	2,13	-1,2	-0,48	-0,45
	MCP/PATH	-0,24	0,75	-4,62	-1,07	-0,15	-0,59	1,99	-0,91	0	0
CZ	conopt	-0,25	0,66	-8,93	-2,61	0,18	-0,47	2,18	-1,1	-0,39	-0,37
	minos	-0,1*	0*	-0,23*	0,08*	0,07*	-2687448 *	5396070 *	-92219,5 *	181,53 *	174,74 *
	snopt	-0,83	0,44	0	-0,28	-0,96	-0,47	2,18	-1,1	-0,39	-0,37
	CNS/PATH	-0,15	0,7	-10,03	-2,88	0,35	0 *	0 *	0 *	0 *	0 *
	MCP/PATH	-0,1	0,72	-10,5	-2,99	0,42	0 *	0 *	0 *	0 *	0 *
DE	conopt	0,2	0,8	-6,22	-1,86	0,66	-0,51	2,17	-0,85	-0,38	-0,32
	minos	-0,6	0,52	8,18	2,7	-1,38	-99,15 *	2,1E+08 *	-2,5E+08 *	105,74 *	87,98 *
	snopt	-0,03	0,38	-0,15	0,07	0,05	-0,51	2,17	-0,85	-0,38	-0,32
	CNS/PATH	0,01	0,73	-3,04	-0,91	0,2	-0,51	2,17	-0,85	-0,38	-0,32
	MCP/PATH	0,01	0,73	-3,02	-0,91	0,19	-0,51	2,17	-0,85	-0,38	-0,32
DK	conopt	-0,08	0,65	-3,34	-0,93	0,1	-0,86	2,3	-1,36	-0,36	-0,32
	minos	0*	0,64*	-0,28*	0,02*	0,02*	-0,86	2,3	-1,36	-0,36	-0,32
	snopt	-0,11	0,6	0,07	0,13	-0,12	-0,86	2,3	-1,36	-0,36	-0,32
	CNS/PATH	-0,13	0,57	1,83	0,69	-0,23	-0,86	2,3	-1,36	-0,36	-0,32
	MCP/PATH	0,11	0,83	-14,75	-4,33	0,87	-0,86	2,3	-1,36	-0,36	-0,32
EE	conopt	-0,2	0,53	-0,4	-0,14	-0,17	-0,5	2,33	-0,41	-0,25	-0,29
	minos	-0,07	0,56	0,36	0,18	-0,07	103518,9 *	1150301 *	-776589 *	0,06 *	-100 *
	snopt	-0,07	0,56	0,36	0,18	-0,07	-0,5	2,33	-0,41	-0,25	-0,29
	CNS/PATH	-0,2	0,53	-0,41	-0,14	-0,18	-0,5	2,33	-0,41	-0,25	-0,29
	MCP/PATH	-0,2	0,53	-0,41	-0,14	-0,18	-0,5	2,33	-0,41	-0,25	-0,29
ES	conopt	0,59	0,79	2,49	1,9	0,39	-0,46	2,39	-0,63	-0,22	-0,3
	minos	-0,02	0,54	0	0,18	-0,05	-917575 *	-99,9 *	1867309 *	-72,09 *	-100 *
	snopt	-0,03	0,54	-0,01	0,17	-0,05	-0,46	2,39	-0,63	-0,22	-0,3
	CNS/PATH	1,74	1,25	7,06	5,07	1,19	-0,46	2,39	-0,63	-0,22	-0,3
	MCP/PATH	1,74	1,25	7,07	5,07	1,19	-0,46	2,39	-0,63	-0,22	-0,3
FI	conopt	-0,02	0,45	-0,73	-0,19	0,07	-0,68	2,53	-1,28	-0,43	-0,39
	minos	-0,05	0,44	-0,42	-0,08	-0,01	-100 *	40,88 *	-98,01 *	-88,66 *	-80,03 *
	snopt	-0,05	0,44	-0,36	-0,06	-0,03	-0,68	2,53	-1,28	-0,43	-0,39
	CNS/PATH	-0,21	0,41	1,41	0,59	-0,49	-0,68	2,53	-1,28	-0,43	-0,39
	MCP/PATH	-0,21	0,41	1,41	0,59	-0,49	-0,68	2,53	-1,28	-0,43	-0,39
FR	conopt	-0,31	1,04	-1,19	-0,35	-0,35	-0,45	1,68	-0,89	-0,33	-0,34
	minos	-0,12	1,06	1,17	1,03	-0,51	-0,45	1,68	-0,89	-0,33	-0,34
	snopt	-0,22	1,05	-0,07	0,3	-0,43	-0,45	1,68	-0,89	-0,33	-0,34
	CNS/PATH	-0,66	1,01	-6,14	-3,14	0,02	-0,45	1,68	-0,89	-0,33	-0,34
	MCP/PATH	-0,29	1,04	-0,93	-0,2	-0,37	-0,45	1,68	-0,89	-0,33	-0,34
GR	conopt	0,03	0,89	-0,01	0,34	-0,01	0 *	0 *	0 *	0 *	0 *
	minos	-0,02	0,88	-0,24	0,2	-0,02	-100 *	-100 *	-100 *	-55,81 *	-100 *

	snopt	0	0,88	-0,14	0,26	-0,02	0 *	-100 *	89,17 *	0 *	0 *
	CNS/PATH	0,42	0,92	1,81	1,44	0,04	-0,27	1,8	-0,6	-0,15	-0,27
	MCP/PATH	0,42	0,92	1,81	1,44	0,04	-0,27	1,9	-0,71	-0,14	-0,26
HU	conopt	-0,48	0,22	-3,69	-1,17	-0,24	-0,72	2,53	-0,95	-0,39	-0,42
	minos	-0,16 *	0,32 *	0 *	0,12 *	0,01 *	-20,52 *	-244,35 *	-4,18 *	-94,7 *	-100 *
	snopt	0	0	0	0	0	0 *	-100 *	98,85 *	0 *	0 *
	CNS/PATH	0	0,43	1,9	0,56	-0,16	0 *	0 *	0 *	0 *	0 *
	MCP/PATH	0 *	0,43 *	1,9 *	0,56 *	-0,16 *	0 *	0 *	0 *	0 *	0 *
IE	conopt	0,01	0,58	-0,91	-0,17	0,11	-0,69 *	2,31 *	-0,34 *	-0,1 *	-0,09 *
	minos	-0,05	0,57	-0,28	-0,01	0,01	19,95 *	54,26 *	-345,72 *	-70,58 *	-60,16 *
	snopt	-0,06	0,57	-0,23	0,01	0	-0,69 *	2,31 *	-0,34 *	-0,1 *	-0,09 *
	CNS/PATH	-0,33	0,55	2,64	0,75	-0,49	-0,71	2,27	-0,28	0	0
	MCP/PATH	-0,33	0,55	2,64	0,75	-0,49	0 *	-1,76 *	1,07 *	0,1 *	0,08 *
LT	conopt	1,26	1,92	-1,97	0,51	-1,19	-0,19 *	17,95 *	-12,42 *	-2,65 *	-3,04 *
	minos	0,02	1,86	-10,27	-2,5	-1,5	-77,85 *	-251,11 *	-47,27 *	-19,06 *	-100 *
	snopt	0,7	1,89	-5,68	-0,84	-1,33	-	-	-	-	-
	CNS/PATH	2,05	1,96	3,11	2,36	-0,98	0 *	-1,5 *	0,34 *	-0,13 *	-0,15 *
	MCP/PATH	2,05	1,96	3,11	2,36	-0,98	0,09 *	-0,26 *	-0,17 *	0,92 *	1,06 *
NL	conopt	1,22 *	-1,65 *	-482,01 *	UNDF *	22,92 *	-	-	-	-	-
	minos	0,88 *	-1,71 *	-477,94 *	UNDF *	22,18 *	-	-	-	-	-
	snopt	-15,49 *	-3,43 *	-284,69 *	UNDF *	-22,55 *	-	-	-	-	-
	CNS/PATH	1,12 *	-1,67 *	-480,79 *	UNDF *	22,7 *	-	-	-	-	-
	MCP/PATH	1,12 *	-1,67 *	-480,79 *	UNDF *	22,7 *	-	-	-	-	-
NO	conopt	-0,19	0,33	0,74	0,33	-0,26	-0,74	2,6	-1,26	-0,28	-0,18
	minos	-0,04	0,37	-0,21	-0,01	-0,02	-100 *	-48,42 *	-233,24 *	-88,23 *	-55,69 *
	snopt	-0,02	0,37	-0,3	-0,04	0	-0,74	2,6	-1,26	-0,28	-0,18
	CNS/PATH	-0,08	0,36	0,05	0,08	-0,09	-0,74	2,6	-1,26	-0,28	-0,18
	MCP/PATH	-0,08	0,36	0,05	0,08	-0,09	-0,73 *	2,6 *	-1,27 *	-0,28 *	-0,18 *
PL	conopt	-0,37	0,65	-0,87	-0,26	-0,25	-0,46	2,04	-0,58	-0,23	-0,23
	minos	-0,15	0,68	1,38	0,74	-0,22	-95,87 *	-1,2E+09 *	1,2E+09 *	56,28 *	58,19 *
	snopt	-0,18	0,68	1,08	0,6	-0,22	-0,46	2,04	-0,58	-0,23	-0,23
	CNS/PATH	-0,2	0,68	0,95	0,55	-0,22	-0,46	2,04	-0,58	-0,23	-0,23
	MCP/PATH	-0,2	0,68	0,95	0,55	-0,22	-0,46	2,04	-0,58	-0,23	-0,23
PT	conopt	0,01	0,63	0,17	0,27	-0,08	-0,5 *	2,29 *	-0,92 *	-0,36 *	-0,53 *
	minos	-0,02	0,62	0,07	0,21	-0,1	-102,69 *	-84,58 *	-63,86 *	-34,82 *	-50,68 *
	snopt	-0,02	0,62	0,06	0,21	-0,1	-0,5	2,29	-0,92	-0,36	-0,53
	CNS/PATH	11,08	3,57	41,6	23,27	8,82	0 *	0 *	0 *	0,17 *	0,25 *
	MCP/PATH	4,62	1,9	17,43	10,06	3,78	-0,5	2,29	-0,92	-0,36	-0,53
RO	conopt	0,1	1,08	1,77	1,17	-0,08	-0,3 *	1,68 *	-0,45 *	-0,15 *	-0,2 *
	minos	-0,14	1,02	0,55	0,56	-0,15	14104,15 *	18439,12 *	-54564,1 *	-76,45 *	-100 *
	snopt	-0,15	1,02	0,53	0,54	-0,15	-0,3	1,68	-0,45	-0,15	-0,2
	CNS/PATH	0,69	1,21	4,71	2,68	0,08	-0,29	1,51	-0,34	-0,08	-0,1
	MCP/PATH	0,69	1,21	4,71	2,68	0,08	-0,28	1,55	-0,39	0,07	0,09
SE	conopt	0,02	1,25	-12,93	-2,05	-0,18	-0,62 *	1,71 *	-0,92 *	0,46 *	0,37 *
	minos	-0,21	1,17	-8,1	-0,91	-0,79	-3E+09 *	-1,4E+10 *	3,5E+09 *	-123,14 *	-100 *
	snopt	-0,17	0,04	-2,9	0,28	-1,09	0,21	2,43	0,1	0,57	0,46
	CNS/PATH	-0,12	1,2	-9,96	-1,36	-0,56	0 *	0 *	-0,47 *	0,16 *	0,13 *
	MCP/PATH	0,59	1,46	-24,23	-4,53	1,29	0,02 *	0,25 *	-0,44 *	-0,03 *	-0,03 *
SK	conopt	0	0,98	4,69	1,61	-0,19	-0,37 *	1,75 *	-0,47 *	-0,2 *	-0,21 *
	minos	-0,3	0,89	3,12	1,01	-0,41	-5,88 *	22,22 *	-32,98 *	-136,47 *	-144,81 *
	snopt	-0,35	0,87	2,86	0,91	-0,45	-0,37	1,75	-0,47	-0,2	-0,21
	CNS/PATH	1,51	1,46	12,48	4,58	0,89	-0,37	1,75	-0,47	-0,2	-0,21
	MCP/PATH	1,51	1,46	12,48	4,58	0,89	-0,37	1,75	-0,47	-0,2	-0,21
SL	conopt	0	0,83	4,73	1,72	-0,37	-0,57 *	2,04 *	-0,41 *	-0,26 *	-0,27 *
	minos	-0,3	0,74	0,46	0,14	-0,39	9,98 *	59,71 *	-54,84 *	-31,09 *	-33,15 *
	snopt	-0,26	0,75	1	0,34	-0,39	-0,57	2,04	-0,41	-0,26	-0,27
	CNS/PATH	-0,67	0,62	-5,53	-2,05	-0,38	-0,57	2,04	-0,41	-0,26	-0,27
	MCP/PATH	-0,66	0,63	-5,3	-1,96	-0,38	-0,52	2,06	-0,53	-0,26	-0,28
TR	conopt	-0,27	0,55	-0,13	0,2	-0,2	-0,39	2,12	-0,08	-0,12	-0,16
	minos	0 *	-0,46 *	3,65 *	2,85 *	-0,14 *	9929254 *	-2,8E+07 *	0 *	-69,12 *	30,91 *
	snopt	0	0	0	0	0	0	-100	72,58	0	0
	CNS/PATH	0,14	0,64	1,83	1,51	-0,13	0 *	0 *	-0,03 *	0 *	0 *
	MCP/PATH	0,07 *	0,63 *	1,38 *	1,22 *	-0,13 *	0 *	0 *	0 *	0 *	0 *
UK	conopt	-0,02	0,93	0,21	0,38	-0,1	0 *	0 *	0 *	0 *	0 *
	minos	-0,08	0,93	-0,39	0,09	-0,11	-100 *	-100 *	-100 *	-86,45 *	-100 *
	snopt	-0,04	0,93	-0,01	0,27	-0,1	0 *	-100 *	132,58 *	0 *	0 *
	CNS/PATH	0,26	0,96	3,25	1,86	-0,03	0 *	0 *	0 *	0 *	0 *
	MCP/PATH	0,23	0,96	3,01	1,74	-0,03	-0,84	1,63	0,89	0	0

4 Conclusion

The results indicated a higher sensitivity of the model VŠEMOR2 to the selected solving method and by NLP also to individual solvers. For model STDCGE it was a little different. The model itself is not very sensitive to the choice of solution method, but rarely solution leads to good result. In other words, if we already get a result, differences between solvers are minimal. Interesting results were differences between the solutions obtained for different solvers in solving the model VŠEMOR2 through NLP. The results look like individual solvers have held different economic theories.

Another conclusion may be a recommendation for the construction of CGE models. Given the number and especially communicativeness of solvers is good to start with NLP. Solvers are this case giving great feedback so any problem with the model is not difficult to find and fix. The finished model is then advisable to reformulate as MCP or CNS for several reasons. The first and for large models perhaps the most important is the speed of calculation. This is for NLP acceptable for routine analysis, but for quality sensitivity analysis of model that is not it. Here really are differences in the speed of calculation. Another, somewhat alibistic reason to the analyst is lack of doubt about the results, when for NLP solvers gave different but mathematically equally good results. For MCP and CNS it is possible to solve models by several solvers, but they work more complementary, where one

fails, the other can find the result, which was confirmed in our case, when CONOP could not solve model and PATH can.

Comprehensive view of the results indicates more findings. The first may be the relative inadequacy of raw STDCGE model for comparative analysis of selected countries. The model proved to be stable when using multiple solution methods, but only for some countries, thus only for certain combinations of input data in the SAM. VŠEMOR2 model is doing differently. It is able to provide results for almost all countries regardless of the method of calculation. However, the results vary between different solution methods, so that depend on analyst who evaluates the results which are correct. Both models have their advantages and disadvantages and remains on the analyst to which results he lean.

Acknowledgements

This paper was created within the project of Internal Grant Agency IGA F4/24/2014.

References

[1] Beutel, J. et al.: *Eurostat Manual of Supply, Use and Input-Output Tables*. European Commision, Luxembourg, 2008.
[2] Bora, B., Cernat, L. and Turrini, A.: *Duty and Quota Free Market Acces for LDCs: Further Evidence from CGE Modeling*. United Nations, Geneva, 2002.
[3] Burfisher, M. E.: *Introduction to Computable General Equilibrium Models*. Cambridge University Press, 2011.
[4] Cardenete, M. A., Guerra, A. I. and Sancho, F.: *Applied General Equilibrium*. Springer, 2012
[5] Decaluwé, B., et al.: *The PEP Standard Computable General Equilibrium Model*. Single Country, Static Version. Poverty and Economic Policy (PEP) Research Network, 2009.
[6] Decaluwé, B., Martens, A. and Monette, M.: *Macroclosures in Open Economy CGE Models: A Numerical Reappraisal*. Université de Montréal, Montreal, 1987.
[7] Drozdik, J.: Všeobecný algebraický modelovací systém – GAMS. Diplomová práca, FMFI UK, Bratislava, 2006.
[8] Herr, B., Maussner, A.: *Dynamic General Equilibrium Modeling*. Springer, Berlin, 2009.
[9] Hosoe, N., Gasawa, K. and Hashimoto, H.: *Textbook of Computable General Equilibrium Modeling: Programming and Simulations*. Palgrave MacMillan, 2010.
[10] Křístková, Z.: Impact of R&D investment on economic growth of the Czech Republic-A recursively dynamic CGE approach. *Prague Economic Papers*, 2012, 2012.4: 412-433.
[11] McCarl, B. A.: *McCarl Expanded GAMS User Guide Version 23.6.* GAMS Development, Washington, 2011.
[12] Pániková, L.: *Alternatívne uzávery CGE modelov*. Diplomová práca, FMFI UK, Bratislava, 2007.
[13] Školuda, V. Comparison of V4 economies using CGE model. In: *Mathematical Methods in Economics 2013*. Jihlava: College of Polytechnics Jihlava, 2013, s. 808-813.
[14] Školuda, V. Kvantifikácia dopadov vybraných hospodárskych opatrení pomocou CGE modelu Slovenskej republiky. Praha 10.12.2012 – 12.12.2012. In: *Mezinárodní vědecký seminář – Nové trendy v ekonometrii a operačním výzkumu*. Bratislava: EKONÓM, 2012, s. 1–9.
[15] Školuda, V. Possibilities of computable general equilibrium techniques for analysis of the impact of select government policies using the CGE model of the Czech Republic. Karviná. In: *Mathematical Methods in Economics 2012.* Opava : University in Opava, 2012, s. 879–884.
[16] Školuda, V. Využitie ekonometrických odhadov pri kalibrácií CGE modelov. Praha 14.02.2013. In: *Vědecká konference doktorského studia FIS 2013*. Praha : Oeconomica, 2013, s. 162–173.
[17] Školuda, V; Princ P.: Demonstration of the use of CGE model for comparative analysis of European economies. Paris 29.10.2013 – 31.10.2013. In: *Communication in Innovation and Academic Globalization.*
[18] Szomolányi, K., Lukáčik, M. and Lukáčiková, A.: Estimation of the production function parameters in V4 economies. In: *Mathematical Methods in Economics 2013*. Jihlava: College of Polytechnics Jihlava, 2013, s. 898–903.
[19] Szomolányi, K., Lukáčik, M., and Lukáčiková, A.:The estimate of parameters of production function of Slovak economy: Econometric analysis of nominal wages. Bratislava 30.05.2012 – 01.06.2012. In: *Quantitative Methods in Economics*. Bratislava: EKONÓM, 2012, s. 210–215.
[20] Takeda, S; Ban, K.: *Regional Effects of Trade Liberalization in Japan: A CGE Analysis Based on an Interregional Input-Output Table*. Available at SSRN 1886019, 2010.
[21] Vaittinen, R.: *Trade policies and integration – evaluations with CGE-models.* Helsinki School of Eonomics, Helsikni, 2004.
[22] Vojteková, M.: *Analýza zmien dôchodkového systému pomocou CGE modelov.* Diplomová práca, FMFI UK, Bratislava, 2009.

The Variance of Discounted Rewards in Markov Decision Processes: Laurent Expansion and Sensitive Optimality

Karel Sladký[1]

Abstract. In this paper we consider discounted Markov decision processes with finite state space and compact actions spaces. We present formulas for the variance of total expected discounted rewards along with its partial Laurent expansion. This enables to compare the obtained results with similar results for undiscounted models.

Keywords: discrete-time Markov decision chains, variance of total discounted rewards, Laurent expansion, mean-variance optimality

JEL classification: C44
AMS classification: 90C15

1 Introduction

The usual optimization criteria examined in the literature on stochastic dynamic programming, such as a total discounted or mean (average) reward structures, may be quite insufficient to characterize the problem from the point of a decision maker. To this end it may be preferable if not necessary to select more sophisticated criteria that also reflect variability-risk features of the problem. Perhaps the best known approaches stem from the classical work of Markowitz on mean variance selection rules, i.e. we optimize the weighted sum of the expected total (or average) reward and its variance.

In the present paper we restrict attention on unichain models with finite state space. At first we rederive recursive formulas for total undiscounted and discounted reward variance. The heart of this article is the partial Laurent expansion of the variance of discounted rewards and analysis of this behaviour for the discount factor tending to unity.

2 Notation and Preliminaries

In this note, we consider at discrete time points Markov decision process $X = \{X_n, n = 0, 1, \ldots\}$ with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$, and compact set $\mathcal{A}_i = [0, K_i]$ of possible decisions (actions) in state $i \in \mathcal{I}$. Supposing that in state $i \in \mathcal{I}$ action $a \in \mathcal{A}_i$ is chosen, then state j is reached in the next transition with a given probability $p_{ij}(a)$ and one-stage transition reward r_{ij} will be accrued to such transition.

A (Markovian) policy controlling the decision process, $\pi = (f^0, f^1, \ldots)$, is identified by a sequence of decision vectors $\{f^n, n = 0, 1, \ldots\}$ where $f^n \in \mathcal{F} \equiv \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$ for every $n = 0, 1, 2, \ldots$, and $f_i^n \in \mathcal{A}_i$ is the decision (or action) taken at the nth transition if the chain X is in state i. Let $\pi^m = (f^m, f^{m+1}, \ldots)$, hence $\pi = (f^0, f^1, \ldots, f^{m-1}, \pi^m)$, in particular $\pi = (f^0, \pi^1)$. The symbol E_i^π denotes the expectation if $X_0 = i$ and policy $\pi = (f^n)$ is followed, in particular, $\mathrm{E}_i^\pi(X_m = j) = \sum_{i_j \in \mathcal{I}} p_{i,i_1}(f_i^0) \ldots p_{i_{m-1},j}(f_{m-1}^{m-1})$; $\mathrm{P}(X_m = j)$ is the probability that X is in state j at time m.

Policy π which selects at all times the same decision rule, i.e. $\pi \sim (f)$, is called stationary, hence X is a homogeneous Markov chain with transition probability matrix $P(f)$ whose ij-th element equals $p_{ij}(f_i)$; $\mathrm{E}_i^\pi(X_m = j) = [P^m(f)]_{ij}$ (symbol $[A]_{ij}$ denotes the ij-th element of the matrix A) and $r_i(f_i) := \sum_{j \in \mathcal{I}} p_{ij}(f_i) r_{ij}$ is the expected reward obtained in state i. Similarly, $r(f)$ is an N-column vector of

[1]Institute of Information Theory and Automation of the AS CR, Department of Econometrics, Praha 8, Pod vodárenskou věží 4, e-mail: sladky@utia.cas.cz

one-stage rewards whose i-the elements equals $r_i(f_i)$. The symbol I denotes an identity matrix and e is reserved for a unit column vector.

Recall that (the Cesaro limit of $P(f)$) $P^*(f) := \lim_{n\to\infty} \frac{1}{n}\sum_{k=}^{n-1} P^k(f)$ (with elements $p^*_{ij}(f)$) exists, and if $P(f)$ is aperiodic then even $P^*(f) = \lim_{k\to\infty} P^k(f)$ and the convergence is geometrical. Moreover, if $P(f)$ is unichain, i.e. $P(f)$ contains a single class of recurrent states, then $p^*_{ij}(f) = p^*_j(f)$, i.e. limiting distribution is independent of the starting state.

It is well-known (cf. e.g. [1], [6]) that also $Z(f)$ (fundamental matrix of $P(f)$), and $H(f)$ (the deviation matrix) exist, where

$$Z(f) := (I - P(f) + P^*(f))^{-1}, \quad H(f) := Z(f)\,(I - P^*(f)) \quad \text{satisfy}$$

$$P^*(f)Z(f) = Z(f)P^*(f) = P^*(f), \quad H(f) = Z(f) - P^*(f)$$

$$P^*(f)H(f) = H(f)P^*(f) = 0, \quad (I - P(f))\,Z(f) = Z(f)\,(I - P(f)) = I - P^*(f).$$

As it is known from the literature (see e.g. [3],[12]), when λ is not an eigenvalue of $P(f)$, there exists $\mathrm{R}(\lambda, P(f)) = (\lambda I - P(f))^{-1}$ (called the resolvent of $P(f)$) and for λ sufficiently close to 1 $\mathrm{R}(\lambda, P(f))$ has the following Laurent series expansion:

$$\mathrm{R}(\lambda, P(f)) = \frac{1}{\lambda - 1}P^*(f) + H(f) + \sum_{k=1}^{\infty}(-1)^k(\lambda - 1)^k H^{k+1}(f). \tag{1}$$

Observe that the infinite series on the RHS of (1) converges if $|\lambda - 1| < 1/\rho(H(f))$ where $\rho(H(f))$ denotes the spectral radius of $H(f)$.

Furthermore, if $P(\bar{f})$ is another $N \times N$ transition probability matrix then for any λ sufficiently close to 1, the so-called second Laurent expansion holds

$$\mathrm{R}(\lambda, P(\bar{f})) - \mathrm{R}(\lambda, P(f)) = \mathrm{R}(\lambda, P(\bar{f}))\,(P(\bar{f}) - P(f))\,\mathrm{R}(\lambda, P(f)). \tag{2}$$

3 Reward Variance Under Stationary Policies

Let $\xi^\alpha_n(\pi) = \sum_{k=0}^{n-1}\alpha^k r_{X_k,X_{k+1}}$ with $\alpha \in (0,1)$, resp. $\xi_n(\pi) = \sum_{k=0}^{n-1} r_{X_k,X_{k+1}}$, be the stream of α-discounted, resp. undiscounted, rewards received in the n next transitions of the considered Markov chain X if policy $\pi = (f^n)$ is followed. Supposing that $X_0 = i$, on taking expectation we get for the first and second moments of $\xi^\alpha_n(\pi)$,

$$v_i^{\alpha(1)}(\pi, n) := \mathrm{E}^\pi_i(\xi^\alpha_n(\pi)) = \mathrm{E}^\pi_i \sum_{k=0}^{n-1}\alpha^k r_{X_k,X_{k+1}}, \quad v_i^{\alpha(2)}(\pi, n) := \mathrm{E}^\pi_i(\xi^\alpha_n(\pi))^2 = \mathrm{E}^\pi_i(\sum_{k=0}^{n-1}\alpha^k r_{X_k,X_{k+1}})^2. \tag{3}$$

If policy $\pi \sim (f)$ is stationary, the process X is time homogeneous and for $m < n$ we write for undiscounted, resp. α-discounted, random reward $\xi_n = \xi_m + \xi_{n-m}$, resp. $\xi^\alpha_n = \xi^\alpha_m + \alpha^m\xi^\alpha_{n-m}$ (here we delete the symbol π and tacitly assume that $\mathrm{P}(X_m = j)$ and ξ_{n-m} starts in state j). Hence $[\xi^\alpha_n]^2 = [\xi^\alpha_m]^2 + \alpha^{2m}\cdot[\xi^\alpha_{n-m}]^2 + 2\cdot\alpha^m\cdot\xi^\alpha_m\cdot\xi^\alpha_{n-m}$. Then for $n > m$ we can conclude that

$$\mathrm{E}^\pi_i[\xi^\alpha_n] = \mathrm{E}^\pi_i[\xi^\alpha_m] + \alpha^m\mathrm{E}^\pi_i\big\{\sum_{j\in\mathcal{I}}\mathrm{P}(X_m = j)\cdot\mathrm{E}^\pi_j[\xi^\alpha_{n-m}]\big\}. \tag{4}$$

$$\mathrm{E}^\pi_i[\xi^\alpha_n]^2 = \mathrm{E}^\pi_i[\xi^\alpha_m]^2 + \alpha^{2m}\mathrm{E}^\pi_i\big\{\sum_{j\in\mathcal{I}}\mathrm{P}(X_m = j)\cdot\mathrm{E}^\pi_j[\xi^\alpha_{n-m}]^2\big\}$$
$$+2\cdot\alpha^m\cdot\mathrm{E}^\pi_i[\xi^\alpha_m]\sum_{j\in\mathcal{I}}\mathrm{P}(X_m = j)\cdot\mathrm{E}^\pi_j[\xi^\alpha_{n-m}]. \tag{5}$$

Using the more appealing notation introduced in (3), from (4) and (5) we conclude for $m = 1$

$$v_i^{\alpha(1)}(f, n+1) = r_i^{(1)}(f_i) + \alpha\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot v_j^{\alpha(1)}(f, n) \tag{6}$$

$$v_i^{\alpha(2)}(f, n+1) = r_i^{(2)}(f_i) + 2\cdot\alpha\cdot\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\ r_{ij}\cdot v_j^{\alpha(1)}(f, n) + \alpha^2\cdot\sum_{j\in\mathcal{I}} p_{ij}(f_i)\, v_j^{\alpha(2)}(f, n) \tag{7}$$

where $r_i^{(1)}(f_i) := \sum_{j\in\mathcal{I}} p_{ij}(f_i)\ r_{ij}, \quad r_i^{(2)}(f_i) := \sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij}]^2$.

Since the variance $\sigma_i^\alpha(f,n) = v_i^{\alpha(2)}(f,n) - [v_i^{\alpha(1)}(f,n)]^2$ from (6),(7) we get

$$\begin{aligned}\sigma_i^\alpha(f,n+1) &= r_i^{(2)}(f_i) + \alpha^2 \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^\alpha(f,n) + 2\cdot\alpha\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\ r_{ij}\cdot v_j^{\alpha(1)}(f,n)\\ &\quad -[v_i^{\alpha(1)}(f,n+1)]^2 + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)[v_j^{\alpha(1)}(f,n)]^2 \qquad (8)\\ &= \sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij}+\alpha\cdot v_j^{\alpha(1)}(f,n)]^2 - [v_i^{\alpha(1)}(f,n+1)]^2 + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^\alpha(f,n). \qquad (9)\end{aligned}$$

Using matrix notations equations (6), (7) can be written as:

$$v^{\alpha(1)}(f,n+1) = r^{(1)}(f) + \alpha P(f)\cdot v^{\alpha(1)}(f,n) \qquad (10)$$

$$v^{\alpha(2)}(f,n+1) = r^{(2)}(f) + 2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f,n) + \alpha^2\cdot P(f)\, v^{\alpha(2)}(f,n) \qquad (11)$$

where $R = [r_{ij}]_{i,j}$ is an $N\times N$-matrix, and
$r^{(2)}(f) = [\ r_i^{(2)}(f_i)]$, $v^{\alpha(2)}(f,n) = [v_i^{\alpha(2)}(f,n)]$, $v^{\alpha(1)}(f,n) = [(v_i^{\alpha(1)}(f,n))^2]$ are column vectors.
The symbol $\circ$ is used for Hadamard (entrywise) product of matrices.

On iterating (10) we easily conclude that there exists $v^{\alpha(1)}(f) := \lim_{n\to\infty} v^{\alpha(1)}(f,n)$ such that

$$v^{\alpha(1)}(f) = r^{(1)}(f) + \alpha P(f)\cdot v^{\alpha(1)}(f) \Longleftrightarrow v^{\alpha(1)}(f) = [I-\alpha P(f)]^{-1} r^{(1)}(f). \qquad (12)$$

Finally, for discounted models on letting $n\to\infty$ there also exists $v^{\alpha(2)}(f) = \lim_{n\to\infty} v^{\alpha(2)}(f,n)$ and by (11)

$$v^{\alpha(2)}(f) = r^{(2)}(f) + 2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f) + \alpha^2\cdot P(f)\, v^{\alpha(2)}(f), \qquad (13)$$

hence

$$v^{\alpha(2)}(f) = [I-\alpha^2\cdot P(f)]^{-1}\left\{\ r^{(2)}(f) + 2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f)\right\}. \qquad (14)$$

On letting $n\to\infty$ from (8), (9) we get for $\sigma_i^\alpha(f) := \lim_{n\to\infty}\sigma_i^\alpha(f,n)$

$$\begin{aligned}\sigma_i^\alpha(f) &= r_i^{(2)}(f_i) + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^\alpha(f) + 2\cdot\alpha\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\ r_{ij}\cdot v_j^{\alpha(1)}(f)\\ &\quad -[v_i^{\alpha(1)}(f)]^2 + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)[v_j^{\alpha(1)}(f)]^2 \qquad (15)\\ &= \sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij}+\alpha\cdot v_j^{\alpha(1)}(f)]^2 - [v_i^{\alpha(1)}(f)]^2 + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^\alpha(f). \qquad (16)\end{aligned}$$

Hence in matrix notation

$$\sigma^\alpha(f) = r^{(2)}(f) + \alpha^2\cdot P(f)\cdot\sigma^\alpha(f) + 2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f) - [v^{\alpha(1)}(f)]^2 + \alpha^2\cdot P(f)\cdot[v^{\alpha(1)}(f)]^2. \qquad (17)$$

After some algebra (17) can be also written as

$$\sigma^\alpha(f) = [I-\alpha^2\cdot P(f)]^{-1}\cdot\{\ r^{(2)}(f) + 2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f) - [v^{\alpha(1)}(f)]^2\}. \qquad (18)$$

(18) is similar to the formula for the variance of discounted rewards obtained by Sobel [11] by different methods.

4 Laurent Expansions of Discounted Variance

To begin with, first observe that for $\alpha := \lambda^{-1}$ we have $(\lambda I - P(f))^{-1} = \alpha(I-\alpha P(f))^{-1}$ and (1) takes on the form

$$\alpha(I-\alpha P(f))^{-1} = \frac{\alpha}{1-\alpha}P^*(f) + H(f) + \sum_{k=1}^{\infty}(-1)^k\left(\frac{1-\alpha}{\alpha}\right)^k H^{k+1}(f) \qquad (19)$$

Introducing $\rho := \frac{1-\alpha}{\alpha} \Leftrightarrow \alpha = \frac{1}{1+\rho}$ from (19) we get

$$\alpha(I - \alpha P(f))^{-1} = \rho^{-1}P^*(f) + H(f) + \sum_{k=1}^{\infty}(-1)^k \rho^k H^{k+1}(f). \tag{20}$$

For what follows we shall also need Laurent expansion of $(I - \alpha^2 P(f))^{-1}$. To this end, let $\beta := \alpha^2$ and $\bar{\rho} := \frac{1-\beta}{\beta} \Leftrightarrow \beta = \frac{1}{1+\bar{\rho}}$. Then, in analogy of (19) Laurent expansion of $(I - \beta P(f))^{-1}$ takes on the form

$$\beta(I - \beta P(f))^{-1} = \frac{\beta}{1-\beta}P^*(f) + H(f) + \sum_{k=1}^{\infty}(-1)^k \left(\frac{1-\beta}{\beta}\right)^k H^{k+1}(f) \tag{21}$$

$$(I - \beta P(f))^{-1} = \frac{(\bar{\rho}+1)}{\bar{\rho}}P^*(f) + (\bar{\rho}+1)H(f) + (\bar{\rho}+1)\sum_{k=1}^{\infty}(-1)^k \bar{\rho}^k H^{k+1}(f) \tag{22}$$

Assumption A. There exists state $i_0 \in \mathcal{I}$ that is accessible from any state $i \in \mathcal{I}$ for every $f \in \mathcal{F}$, i.e. for every $f \in \mathcal{F}$ the transition probability matrix $P(f)$ is *unichain*.

Lemma 3.1. If Assumption A holds then

$$\alpha\, v^{\alpha(1)}(f) = \rho^{-1}\, \bar{g}^{(1)}(f) \cdot e + w^{(1)}(f) + \sum_{k=1}^{\infty}(-\rho)^k\, w^{(k,1)}(f) \tag{23}$$

where $\bar{g}^{(1)}(f) = p^*(f) \cdot r^{(1)}(f)$, $w^{(1)}(f) = H(f) \cdot r^{(1)}(f)$, and $w^{(k,1)}(f) = H^{k+1}(f) \cdot r^{(1)}(f)$, for $k = 1, 2, \ldots$. In particular, for the ith element of $v^{\alpha(1)}(f)$ it holds

$$\alpha v_i^{\alpha(1)}(f) = \rho^{-1}\bar{g}^{(1)}(f) + w_i^{(1)}(f) + \sum_{k=1}^{\infty}(-\rho)^k\, w_i^{(k,1)}(f), \quad \text{hence}$$

$$v_i^{\alpha(1)}(f) = (1+\rho)\left[\rho^{-1}\bar{g}^{(1)}(f) + w_i^{(1)}(f) - \rho w_i^{(1,1)}(f) + \rho^2 w_i^{(2,1)}(f) + o(\rho^2)\right] \tag{24}$$

where $\lim_{\rho \downarrow 0} o(\rho^2) = 0$.

In what follows we construct partial Laurent expansion of discounted variance $\sigma_i^{\alpha}(f)$. To this end from (16),(17) we conclude that

$$\sigma_i^{\alpha}(f) - \alpha^2 \sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot \sigma_j^{\alpha}(f) = \sum_{j \in \mathcal{I}} p_{ij}(f_i)[r_{ij} + \alpha \cdot v_j^{\alpha(1)}(f)]^2 - [v_i^{\alpha(1)}(f)]^2 \tag{25}$$

To simplify the RHS of (25) the following facts will be extremely useful. Observe that (26), (27) follow from (24), and (28) follows from (24), (26) and (27) after some algebraic manipulations.

Lemma 3.2. If Assumption A holds then

$$\begin{aligned}\alpha^2[v_i^{\alpha(1)}(f)]^2 = {} & [\rho^{-1} \cdot \bar{g}^{(1)}(f) + w_i^{(1)}(f)]^2 - 2 \cdot \bar{g}^{(1)}(f) \cdot w_i^{(1,1)}(f) \\ & + 2\rho \cdot [\bar{g}^{(1)}(f) \cdot w_i^{(2,1)}(f) + w_i^{(1)}(f) \cdot w_i^{(1,1)}(f)] + o(\rho^2)\end{aligned} \tag{26}$$

$$\begin{aligned}(v_i^{\alpha(1)}(f))^2 = {} & \rho^{-2}[\bar{g}^{(1)}(f)]^2 + 2\rho^{-1}\bar{g}^{(1)}(f)[\bar{g}^{(1)}(f) + w_i^{(1)}(f)] + [\bar{g}^{(1)}(f)]^2 + [w_i^{(1)}(f)]^2 + \\ & \bar{g}^{(1)}(f)[4w_i^{(1)}(f) - 2w_i^{(1,1)}(f)] + 2\rho \cdot [\bar{g}^{(1)}(f) \cdot w_i^{(2,1)}(f) - w_i^{(1)}(f) \cdot w_i^{(1,1)}(f) \\ & + [w_i^{(1)}(f)]^2 - 2 \cdot \bar{g}^{(1)}(f) \cdot w_i^{(1,1)}(f) + \bar{g}^{(1)}(f) \cdot w_i^{(1)}(f)] + o(\rho^2)\end{aligned} \tag{27}$$

$$\begin{aligned}[r_{ij} + \alpha v_j^{\alpha(1)}(f)]^2 = {} & [r_{ij}]^2 + 2 \cdot \alpha \cdot v_j^{\alpha(1)}(f) \cdot r_{ij} + \alpha^2[v_j^{\alpha(1)}(f)]^2 \\ = {} & [\rho^{-1}\bar{g}^{(1)}(f) + w_j^{(1)}(f)]^2 - 2 \cdot \bar{g}^{(1)}(f) \cdot w_j^{(1,1)}(f) + [r_{ij}]^2 + 2\rho \cdot [\bar{g}^{(1)}(f) \cdot w_j^{(2,1)}(f) \\ & + w_j^{(1)}(f) \cdot w_j^{(1,1)}(f)] + 2 \cdot r_{ij} \cdot [\rho^{-1}\bar{g}^{(1)}(f) + w_j^{(1)}(f) + \rho w_j^{(1,1)}(f)] + o(\rho^2)\end{aligned} \tag{28}$$

Lemma 3.3. If Assumption A holds then

$$\sum_{j\in\mathcal{I}} p_{ij}(f_i)\left[r_{ij}+\alpha\cdot v_j^{\alpha(1)}(f)\right]^2-[v_i^{\alpha(1)}(f)]^2=\sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij}+w_j]^2-[\bar{g}^{(1)}(f))-w_i]^2+O(\rho)+o(\rho^2) \quad (29)$$

where

$$O(\rho)=2\rho\sum_{j\in\mathcal{I}} p_{ij}(f_i)w_j^{(1,1)}(f)[w_j^{(1)}(f)+r_{ij}]+[w_i^{(1)}(f)+\bar{g}^{(1)}(f)]w_i^{(1,1)}(f)+[w_i^{(1)}(f)-\bar{g}^{(1)}(f)]w_i^{(1)}(f). \quad (30)$$

Proof. In virtue of (27),(28) we can conclude that

$$\begin{aligned}
&\sum_{j\in\mathcal{I}} p_{ij}(f_i)\left[r_{ij}+\alpha\cdot v_j^{\alpha(1)}(f)\right]^2-[v_i^{\alpha(1)}(f)]^2=\sum_{j\in\mathcal{I}} p_{ij}(f_i)\left\{\left[r_{ij}+\alpha\cdot v_j^{\alpha(1)}(f)\right]^2-[v_i^{\alpha(1)}(f)]^2\right\}\\
&=\sum_{j\in\mathcal{I}} p_{ij}(f_i)\left\{[\rho^{-1}\bar{g}^{(1)}(f)+w_j^{(1)}(f)]^2-\ 2\cdot\bar{g}^{(1)}(f)\cdot w_j^{(1,1)}(f)+[r_{ij}]^2\right.\\
&+2\rho\cdot[\bar{g}^{(1)}(f)\cdot w_j^{(2,1)}(f)+w_j^{(1)}(f)\cdot w_j^{(1,1)}(f)]+\ 2\cdot r_{ij}\cdot[\rho^{-1}\bar{g}^{(1)}(f)+w_j^{(1)}(f)+\rho w_j^{(1,1)}(f)]\\
&-\rho^{-2}[\bar{g}^{(1)}(f)]^2-2\rho^{-1}\bar{g}^{(1)}(f)[\bar{g}^{(1)}(f)+w_i^{(1)}(f)]-[\bar{g}^{(1)}(f)]^2-[w_i^{(1)}(f)]^2\\
&-\bar{g}^{(1)}(f)[4w_i^{(1)}(f)-2w_i^{(1,1)}(f)]-2\rho\cdot[\bar{g}^{(1)}(f)\cdot w_i^{(2,1)}(f)-w_i^{(1)}(f)\cdot w_i^{(1,1)}(f)\\
&\left.+[w_i^{(1)}(f)]^2-2\cdot\bar{g}^{(1)}(f)\cdot w_i^{(1,1)}(f)-\bar{g}^{(1)}(f)\cdot w_i^{(1)}(f)]\right\}\ +o(\rho^2) \qquad (31)\\
&\text{Since}\qquad \sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij}+w_j^{(1)}(f)-w_i^{(1)}(f)-\bar{g}^{(1)}(f)]=0\\
&\sum_{j\in\mathcal{I}} p_{ij}(f_i)[-w_i^{(1)}(f)+w_j^{(1,1)}(f)-w_i^{(1,1)}(f)]=0,\quad \sum_{j\in\mathcal{I}} p_{ij}(f_i)[w_i^{(1,1)}(f)+w_j^{(2,1)}(f)-w_i^{(2,1)}(f)]=0
\end{aligned}$$

(29) follows from (31) after some algebra.

From (25),(29),(30) we immediately get the following lemma.

Lemma 3.4. If Assumption A holds then

$$\sigma_i^{\alpha}(f)-\alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^{\alpha}(f)=s_i(f_i)+O(\rho)+o(\rho)\quad\text{for}\quad i=1,2,\ldots,N \qquad (32)$$

where

$$s_i(f_i):=\sum_{j\in\mathcal{I}} p_{ij}(f_i)\left\{[r_{ij}+w_j^{(1)}(f)]^2-[\bar{g}^{(1)}(f)+w_i^{(1)}(f)]^2\right\}+O(\rho)+o(\rho)$$

For what follows it will be useful to rewrite (32) in matrix notation. On introducing column vector $s(f)=[s_i(f_i)]_{i=1,\ldots,N}$ (32) can written as

$$\sigma^{\alpha}(f)=(I-\alpha^2P(f))^{-1}s(f)+O(\rho)+o(\rho) \qquad (33)$$

Next lemma adapts Laurent expansion of $(I-\alpha P(f))^{-1}$ to $(I-\alpha^2P(f))^{-1}$. Recall that $\bar{\rho}:=\frac{1-\alpha^2}{\alpha^2}$.

Lemma 3.5. If Assumption A holds then

$$(I-\alpha^2P(f))^{-1}\quad=\quad\frac{(\bar{\rho}+1)}{\bar{\rho}}P^*(f)+(\bar{\rho}+1)H(f)+o(\bar{\rho}) \qquad (34)$$

Observe that $\alpha\to 1\Rightarrow\bar{\rho}\to 0$. In particular, $\lim_{\alpha\to 1}(1-\alpha^2)(I-\alpha^2P(f))^{-1}=P^*(f)$.

Proof. The proof follows immediately from (22).

From (33), (34) we immediately get

Theorem 3.6. If Assumption A holds then

$$\sigma^{\alpha}(f) = \frac{(\bar{\rho}+1)}{\bar{\rho}} P^*(f) + (\bar{\rho}+1)s(f) + O(\rho) + o(\rho^2). \tag{35}$$

Moreover, from the well-known Tauberian theorems it holds for undiscounted variance obtain after n transitions

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \sigma(f,n) = \lim_{\alpha\to 1}(1-\alpha)\sigma^{\alpha}(f) = P^*(f)s(f).$$

5 Conclusions

We have received formulas for the variance of discounted rewards in Markov decision chains along with it partial Laurent expansions. Attention was focused only on unichain models and initial terms of the corresponding Laurent expansion that also enables to find formulas for mean variances of undiscounted rewards.

Acknowledgements

This research was supported by the Czech Science Foundation under Grant 13-14445S and by CONACyT (Mexico) and AS CR (Czech Republic) under Project 171396.

References

[1] Berman, A. and Plemmons, R. J.: *Nonnegative Matrices in the Mathematical Sciences.* Academic Press, New York 1979.

[2] Feinberg, E.A. and Fei, J.: *Inequalities for variances of total discounted costs.* J. Applied Probability **46** (2009), 1209–1212.

[3] Kato, T.: *Perturbation Theory for Linear Operators.* Springer Verlag, Berlin 1966.

[4] Mandl, P.: *On the variance in controlled Markov chains.* Kybernetika **7** (1971), 1–12.

[5] Miller, B.L. and Veinott, A.F.Jr.: *Discrete dynamic programming with a small interest rate.* Annals Math. Statistics **39** (1968), 366–370.

[6] Puterman, M.L.: *Markov Decision Processes – Discrete Stochastic Dynamic Programming.* Wiley, New York 1994.

[7] Righter, R.: *Stochastic comparison of discounted rewards.* J. Applied Probability **48** (2011), 293–294.

[8] Sladký, K.: *On the set of optimal controls for Markov chains with rewards.* Kybernetika **10** (1974), 526–547.

[9] Sladký, K.: *On mean reward variance in semi-Markov processes.* Math. Methods Oper. Res. **62** (2005), 387–397.

[10] Sladký, K.: *Risk sensitive and risk-neutral optimality in Markov decision chains; a unified approach.* In: Proceedings of the International Scientific Conference Quantitative Methods in Economics (Multiple Criteria Decision Making XVI), M. Reiff, ed., University of Economics, Bratislava 2012, pp. 201–205.

[11] Sobel, M.: *The variance of discounted Markov decision processes.* J. Applied Probability **19** (1982), 794–802.

[12] Veinott, A.F.Jr.: *Discrete dynamic programming with sensitive discount optimality criteria.* Annals Math. Statistics **40** (1969), 1635–1660.

Analysis of Structural Differences between Czech Economy and Euro Area using Impulse Responses

Martin Slanicay[1]

Abstract. The goal of this paper is to examine structural differences between the Czech economy and the euro area. For this purpose I use New Keynesian DSGE model of two economies which I estimate on the data of the Czech economy and the euro area, using Bayesian techniques. The applied method is based on comparison of impulse responses and examines differences in behavior of both economies in responses to macroeconomic shocks. Results suggest that in general responses of domestic variables have the same shape as responses of foreign variables. However, the responses of domestic variables are always larger, more persistent, and more hump-shaped. The main differences are in responses to productivity shocks in tradables and to shocks in government expenditures, while the biggest similarities are in responses to productivity shocks in non-tradables and to shocks in investment efficiency.

Keywords: DSGE model, Bayesian estimation, structural differences, impulse responses.

JEL classification: C51, C68, E32
AMS classification: 91B51

1 Motivation

Structural differences between economies are regarded as one of the main causes of a potential suboptimality of common monetary policy. The reason is that they contribute to asymmetric behavior of economies during the business cycle. Therefore, the goal of this paper is to examine structural differences between the Czech economy and the euro area. For this purpose I use New Keynesian DSGE model of two economies which I estimate on the data of the Czech economy and the euro area, using Bayesian techniques. The applied method is based on comparison of impulse responses and examines differences in behavior of both economies in responses to macroeconomic shocks.

The rest of this paper proceeds as follows. In the next section I describe the employed model in a brief non-technical manner. In the third section I briefly discuss some issues related to the estimation of the model. The fourth section contains the analysis based on comparison of impulse responses. Finally, the last section concludes.

2 Model

I use a New Keynesian DSGE model of two economies, originally presented in Kolasa [4].[1] In this section I restrict my description of this model to a brief non-technical overview of its structure. The model assumes that there are only two economies in the world: a domestic economy (represented by the Czech economy) and a foreign economy (represented by the euro area).

The model assumes five types of representative agents in each economy. Households consume tradable and non-tradable goods produced by firms. There is an assumption of habit formation in consumption and an assumption that consumption of a final tradable good requires consumption of ω units of non-tradable distribution services. Households trade bonds, too, and their intertemporal choice about consumption is influenced by preference shocks. Households supply labor and set wages on a monopolistically competitive

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno 602 00, slanicay@mail.muni.cz

[1]I depart from the original specification of the model in several aspects, however, these modifications are generally minor.

labor market. Their labor supply is influenced by labor supply shocks and their wage-setting is subject to a set of labor demand constraints and to the Calvo constraint on the frequency of wage adjustment, see Calvo [1]. According to the Calvo constraint, in every period each household resets its wage with probability $1-\theta_w$ and keeps its wage unchanged with probability θ_w. Households also accumulate capital, which they rent to firms. Capital accumulation is subject to investment-specific technological shocks and to adjustment costs.

There are two types of firms in each economy: producers of tradable goods and producers of non-tradable goods. Both of them employ a Cobb-Douglas production function with constant returns to scale. Productivity in both sectors is influenced by productivity shocks. Firms hire labor on the labor market and sell their goods on monopolistically competitive goods markets. They set prices on the goods market subject to a set of demand constraints and to the Calvo constraint on the frequency of price adjustment, see Calvo [1]. According to the Calvo constraint, in every period each firm resets its price with probability $1-\theta_p$ and keeps its price unchanged with probability θ_p.

The fiscal authority collects lump-sum taxes, which it uses for government expenditures and transfers to households so that the state budget is balanced each period. Government expenditures consist only of domestic non-tradable goods and are modeled as a stochastic process - a government expenditures shock. Given our assumptions about households, Ricardian equivalence holds in this model. The monetary authority follows a Taylor–type rule, and deviations from this rule are explained as monetary shocks. The model is closed with an assumption of a complete bond market and with an assumption of goods and labor markets clearing.

3 Estimation

For the estimation of the model I used the quarterly data of the Czech economy and the euro area-17 economy from the 1st quarter of 2000 to the 2nd quarter of 2013. The employed data series were downloaded from the web database of Eurostat. I use following 14 time series (seven for each economy): real GDP, consumption, investment, HICP, real wage, short-term interest rate and internal exchange rate defined as prices of non-tradable goods relative to prices of tradable goods. Except for the nominal interest rates, all observed variables are seasonally adjusted and expressed as demeaned 100*log differences. Nominal interest rates are demeaned and expressed as quarterly rates per cent.

The model is estimated with Random Walk Chain Metropolis-Hastings algorithm, see Hastings [3], using Dynare toolbox for Matlab.[2] I generated four independent chains, each with 2,000,000 draws. From each chain I used only 25% percents of last draws, i. e. 1,500,000 initial draws from each chain were discarded. Average acceptance rate in each chain is about 27%, which is in line with the informal recommendation about ideal acceptance rate. According to the MCMC convergence diagnostics, smoothed shocks and variables, and the prior and posterior distributions of the parameters, the model was estimated successfully.[3]

4 Analysis Based on Impulse Responses

Structural differences can be seen as differences in behavior of both economies. I use impulse responses to compare the behavior of both economies in responses to various shocks. Each subfigure displays a response of a domestic variable to a domestic shock (solid line) and a response of a foreign variable to a foreign shock of the same type (dashed line). A horizontal axis represents a timeline with a time unit equals to one quarter. On a vertical axis is measured a deviation of the examined variables from its respective steady state, expressed in percent. I compare responses of domestic and foreign variables to the shocks of the same type, where volatility and persistence of these shocks is given by their estimated posterior means.

[2] http://www.dynare.org/

[3] For the sake of space I do not present these things in the paper. For the same reasons I also do not provide detailed justification for the calibration of several structural parameters and for the prior setting of the estimated parameters, as well as interpretation of the parameters estimates. However, all these things are available on the request from the author.

4.1 Productivity Shocks in Tradables

Figure 1 displays the impulse response functions of selected variables to the productivity shock in tradables. We can see that behavior of both economies is quite different. We can see that the average domestic productivity shock in tradables has larger magnitude and much higher persistence than the average foreign productivity shock in tradables. Given the fact that the Czech economy is a post-communist country which few years ago underwent transformation to the market economy, such larger magnitude and much higher persistence of the domestic productivity shock in tradables is understandable. The nature of the domestic productivity shock in tradables results in much larger and persistent effects on the domestic variables. The responses of the domestic variables display more gradual and hump-shaped pattern than responses of the foreign variables.

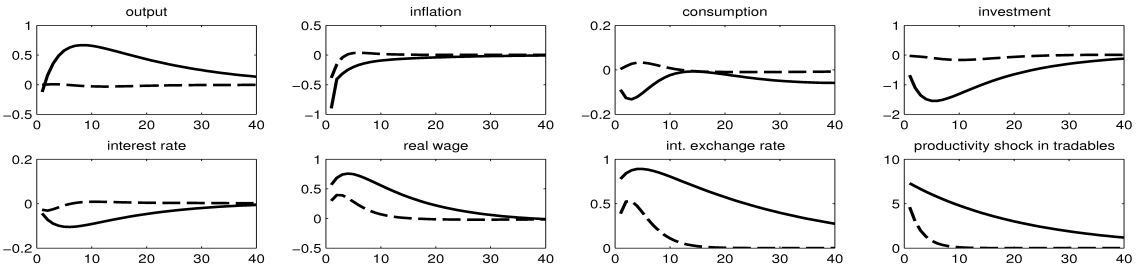


Figure 1 IRFs - productivity shock in tradables

A positive productivity shock in tradables lowers marginal costs in the tradable sector, which in turn lowers inflation and increases output. The decline of inflation induces a reaction of the central bank which lowers nominal interest rate. Real wage rises as a result of sticky nominal wages and the decline in price inflation. The increase of output and real wage should cause an increase of consumption, however, in the case of the domestic economy the decline of inflation is much bigger than the decline of the nominal interest rate which increases real interest rate to the extent that it causes the initial decline of domestic consumption. However, after the initial decline, domestic consumption rises as well as foreign consumption. Because the productivity shock in tradables lowers inflation in the tradable sector more than in the non-tradable sector, it brings about the increase of internal exchange rate. Investment and capital also fall as a result of the increase in the real interest rate and in the internal exchange rate.

4.2 Productivity Shocks in Non-Tradables

On the other hand, responses of the domestic and foreign variables to productivity shock in non-tradables, as depicted in the Figure 2, are quite similar, although the average magnitude of the domestic productivity shock in non-tradables is almost three times larger than the average magnitude of foreign productivity shock in non-tradables. The main difference is in the behavior of consumption and investment such that domestic variables display larger hump than foreign variables.

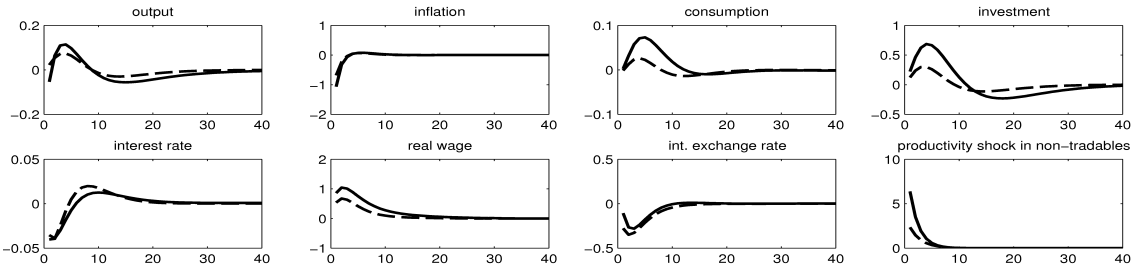


Figure 2 IRFs - productivity shock in non-tradables

A positive productivity shock in non-tradables lowers marginal costs in the non-tradable sector which in turn lowers inflation and increases output. The decline of inflation induces a reaction of the central bank which lowers the nominal interest rate. The real wage rises as a result of sticky nominal wages and the decline in price inflation. The decline of inflation is bigger than the decline of the nominal interest

rate, which causes the real interest rate to increase. However, the increase of output and the real wage dominates the increase of the real interest rate and thus causes an increase of consumption. Because the productivity shock in non-tradables lowers inflation in the tradable sector less than in the non-tradable sector, it results in the decline of the internal exchange rate. Investment and capital rises because the decline of the internal exchange rate dominates the increase of the real interest rate.

4.3 Consumption Preference Shocks

Figure 3 shows reactions of the macroeconomic variables to the consumption preference shock. The average domestic preference shock is more persistent than the average foreign preference shock. It contributes to the fact that the responses of the domestic variables are more persistent and more hump-shaped.

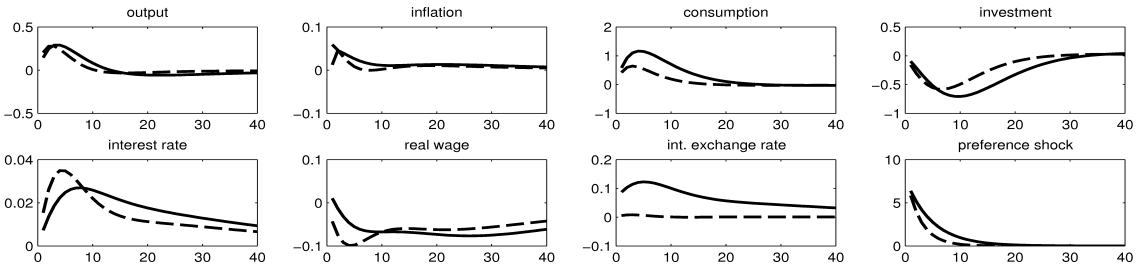


Figure 3 IRFs - consumption preference shock

In the response to the consumption preference shock, consumption rises, which causes an increase of inflation and output and crowding out of investment. The rise of inflation induces a reaction of the central bank which increases nominal interest rate. The real wage falls as a result of sticky nominal wages and the increase in price inflation.

4.4 Labor Supply Shocks

Figure 4 presents responses of the macroeconomic variables to the labor supply shock. The initial impulse given by the average domestic labor supply shock is almost three times larger than in the case of the average foreign labor supply shock. It contributes to the fact that the responses of the domestic variables are larger, more persistent, and more hump-shaped.

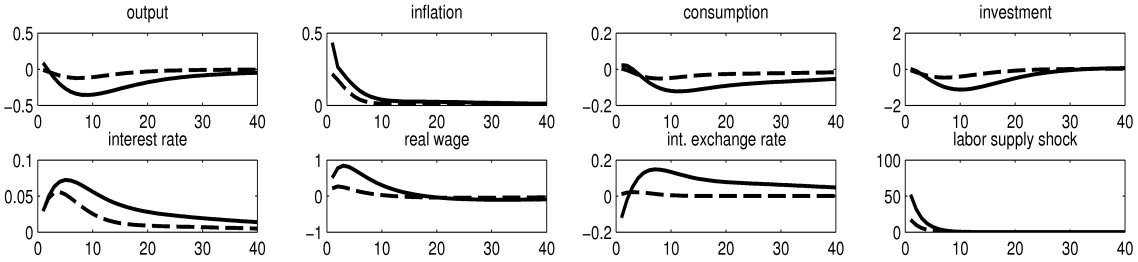


Figure 4 IRFs - labor supply shock

A negative labor supply shock increases the weight of leisure in consumer's utility which causes the labor input to fall, and thus the real wage to rise. The decline of the labor input leads to lower output, consumption, and investment, while the rise of real wage increases the marginal costs in both tradable and non-tradable sector, thus leading to higher inflation. Consequently, higher inflation makes the central bank react and increase the nominal interest rate.

4.5 Shocks in Government Expenditures

In Figure 5 we can see that the average domestic shock in government expenditures is initially almost two times larger than the average foreign shock in government expenditures, and also much more persistent.

Therefore, responses of the domestic variables are larger, more persistent, and more hump-shaped. The main differences are in the behavior of output and investment, while the behavior of inflation is very similar.

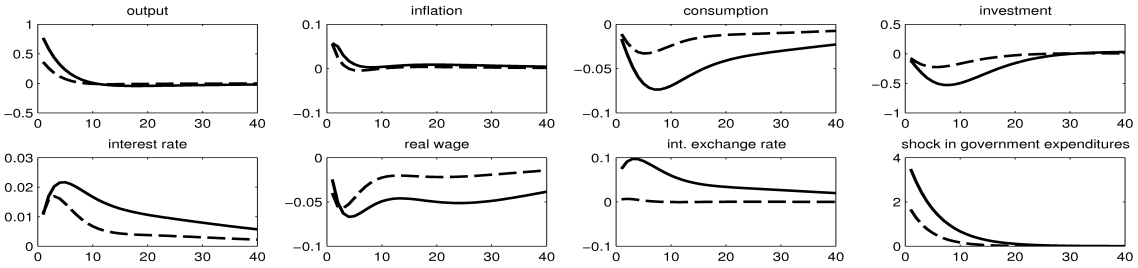


Figure 5 IRFs - shock in government expenditures

The unexpected increase of government expenditures leads to an increase of output and crowding out of consumption and investment.[4] Inflation rises as a consequence of the increase in output, however, only slightly, which in turn induces the reaction of the central bank in the form of the nominal interest rate increase. As a result of sticky nominal wages and of the increase in price inflation, the real wage falls.

4.6 Shocks in Investment Efficiency

As regards responses of both economies to the shock in investment efficiency depicted in Figure 6, we can see that the average domestic shock and the average foreign shock are of the same magnitude and persistence. Also, responses of both domestic and foreign variables are almost identical, however, responses of the domestic variables are a bit more persistent than responses of the foreign variables, as in the previous cases. The main differences are in the responses of output and the internal exchange rate.

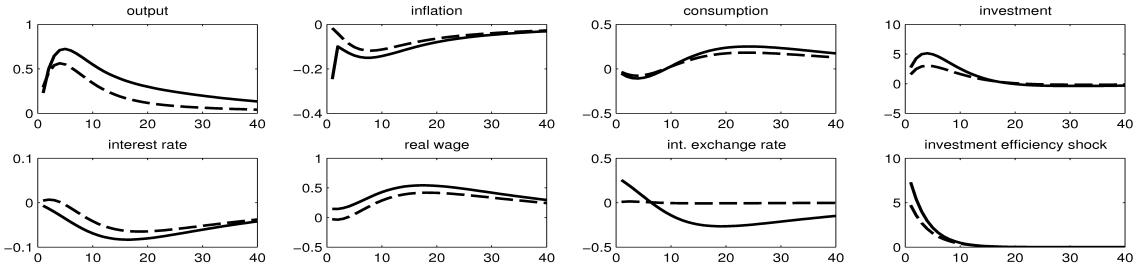


Figure 6 IRFs - shock in investment efficiency

The shock in investment efficiency lowers the price of installed capital, which leads to higher investment and in turn to a higher level of capital. In order to increase investment, households have to initially lower their consumption and increase their labor supply. The increase of labor supply brings about lower marginal costs for the producers, which in turn brings about lower inflation. Central bank responds to lower inflation by lowering its nominal interest rate. As a result of sticky nominal wages and the decline in price inflation, the real wage rises. The slow increase of the real wage leads to the reversal of the development in consumption and the labor input, consumption rises roughly after ten quarters, and the labor input declines.

[4]One can argue that crowding out of consumption is not consistent with the empirical evidence which suggests a positive comovement of consumption and government expenditures. Within DSGE models this effect can be obtained by assuming a large portion of the rule-of-thumb consumers (non-Ricardian households) which behave in non-Ricardian way. These households are always binded by their budget constraints so that they always consume their whole disposable income. For example see Galí, López-Salido and Vallés [2].

4.7 Monetary Policy Shocks

Finally, Figure 7 displays responses of the domestic and foreign variables to the monetary policy shock.[5] We can see that while the effects of monetary policy shocks on development of interest rate and inflation are almost identical, responses of consumption, investment and output show slight differences, with the responses of the domestic variables being larger and more persistent.

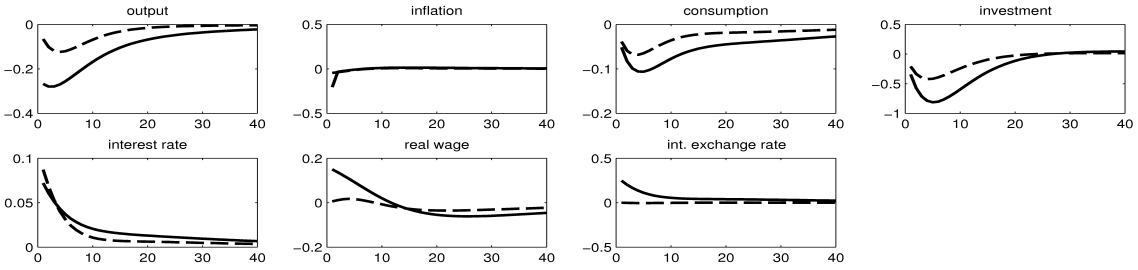


Figure 7 IRFs - monetary policy shock

Monetary policy tightening caused by an unexpected increase of nominal interest rate, leads to decline of inflation. The increase of the nominal interest rate together with the decline of inflation brings about an increase of the real interest rate, which in turn leads to the decline of consumption, investment, and therefore of output. The decline of investment and hence of capital is amplified by the increase in internal exchange rate. As a consequence of sticky nominal wages and the decline in price inflation, real wage rises, which leads to the decline of labor input.

5 Conclusion

In this paper I examined structural differences between the Czech economy and the euro area. For this purpose I used New Keynesian DSGE model of two economies, originally presented in Kolasa [4]. The model is estimated on the data of the Czech economy and the euro area, using Bayesian techniques. The applied method is based on comparison of impulse responses and examines differences in behavior of both economies in responses to shocks. I can say that responses of the domestic variables in general have the same shape as the responses of the foreign variables. However, responses of domestic variables are always larger, more persistent and more hump-shaped. The main differences are in responses to productivity shocks in tradables and to shocks in government expenditures, while the biggest similarities are in responses to productivity shocks in non-tradables and to shocks in investment efficiency.

Acknowledgements

This work was supported by funding of specific research at ESF MU, project MUNI/A/0775/2013.

References

[1] Calvo, G.: Staggered Prices in a Utility Maximizing Framework, *Journal of Monetary Economics* **12(3)** (1983), 383–398.

[2] Galí, J., López-Salido, J. D., and Vallés, J.: Understanding the Effects of Government Spending on Consumption, *Journal of the European Economic Association* **5(1)** (2007), 227–270.

[3] Hastings, W. K.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika* **57(1)** (1970), 97–109.

[4] Kolasa, M.: Structural Heterogeneity or Asymmetric Shocks? Poland and the Euro Area through the Lens of a Two-country DSGE Model, *Economic Modelling* **26(6)** (2009), 1245–1269.

[5] While all other shocks are specified as AR1 processes, monetary policy shocks are described by IID process. Magnitude of the average domestic monetary policy shock is 0.0781, while magnitude of the average foreign monetary policy shock is 0.0964.

Dissimilarity of commodity prices – the results of time series clustering

Sławomir Śmiech[1]

Abstract. The behaviour of commodity prices is difficult to describe due to several reasons. Firstly, there is a large number of different categories of commodities. Secondly, some categories overlap with other categories, while others indirectly compete in the market. Thirdly, although essentially commodity prices react to changes in economic conditions or exchange rates, to a large extent these prices depend on supply disturbances.
The objective of the article is to conduct the classification of the series of commodity prices in the pre-crisis and after-crisis periods and to asses whether the resulting clusters are similar in composition to the commodity indices. The analysis is based on monthly data from the period 1990-01 to 2014-02. All prices and price indices are published by the World Bank. The results obtained in dynamic time warping clustering reveal that commodity price co-movement is more evident in the pre-crisis period. There are only several paths which determine commodity prices.

Key words: *Commodity prices, time series clustering, comovement, dynamic time warping*

JEL Classification: C38, Q02
AMS Classification: 91C20

1. Introduction

Several reasons contribute to difficulties in describing the behaviour of commodity prices. Firstly, there is a large number of different categories of commodities. Secondly, some categories overlap with other categories (for example, biofuel production and energy), while others indirectly compete in the market (for example, the development of one type of crops reduces the supply of other crops cultivated in a given area). Thirdly, although essentially commodity prices react to changes in economic conditions or exchange rates, to a large extent they depend on supply disturbances (such as droughts, floods, armed conflicts, etc). In spite of such complex nature of the behaviour of commodities, the last decade noted their tendency to move together. Frankel [4] argues that the reason for co-movement is the real interest rate, Akram [1] additionally investigates the role of the dollar exchange rates, Svensson [16] discusses the role of shifts in the global supply and demand. Krugman [10] explains the increase in food prices by biofuel production, as biofuel prices are correlated with oil prices. Numerous authors (e.g. Gilbert [5], Phillips and Yu [15], Irwin and Sanders [8]) reason that co-movement is caused by speculations and the existence of price bubbles. From the methodological point of view, the assessment of price co-movement can be conducted with the use of several methods. One of the most common include cointegration (Papież and Śmiech [14], more recently replaced by the panel cointegration approach (Nazlioglu and Soytas [13]), threshold cointegration, (Natanelov et al. [12]) or the general equilibrium model (see e.g. Gohin and Chantert [6]). Other methods incorporate different statistical factor models, e.g FAVAR and PANIC (Byrne et al. [3]).

The objective of the paper is to conduct the classification of the series of commodity prices. The analysis is based on monthly data from three periods: before the global financial crisis, that is the period from 2001-01 to 2008-06, after the crisis, that is the period from 2009-01 to 2014-02, and the period covering the whole sample, that is from 2001-01 to 2014-02. The prices of 54 commodities taken into consideration in the analysis are listed by the World Bank in six categories i.e. energy, metals, beverages, food, raw materials and precious metals. Clustering was conducted with the use of dynamic time warping methods, which allows for the assessment of similarity between series shapes, that is a distance measure which identifies time-shifted patterns among series and seems to be appropriate for the analysis of co-movement of commodities. Eventually, three methods are used to classify time series: Ward's method, complete (hierarchical) and pam (division). The results of the classification are assessed by internal classification measuring the average silhouette width. The clustering conducted provides the answers to the following questions:

- Is moving together of commodity prices similar in intensity in the periods before and after the global financial crisis?
- How many clusters of commodity prices are there and how homogeneous are these clusters?
- Do commodities from the same category (e.g. energy commodities) belong to the same clusters, that is, do their prices behave in a similar manner?
- To what extent do the clusters obtained in the study differ from the indices listed by the World Bank?

[1] Cracow University of Economics, Department of Statistics, Rakowicka 27, Cracow, Poland, e-mail: smiechs@uek.krakow.pl

In comparison to the existing literature, our work differs in one important aspect – the methodology used. Related studies conducted so far assume linear correlations. The methodology used in this study allows us to stretch or compress two time series in order to draw comparisons, which offers a universal analysis of the nonlinear relationship and co-movement of commodity prices.

The rest of the paper is organised as follows. Section 2 describes methodology. Empirical results are discussed in Section 3, and the conclusion is presented in the last section.

2. Methodology

Following the division suggested by Liao [11], three major time series clustering approaches include: raw data approaches, feature-based approaches and model-based approaches. The first ones deal with raw data in the time and frequency domain. They imply working with high dimensional space and are not effective if the raw data are highly noisy. In feature-based approaches certain features are extracted first to be clustered next. Model-based approaches assume that each time series is generated by a particular time series model. To obtain dissimilarity between series, models are fitted and then discrepancies between them are looked for. The disadvantage of the feature-based and model-based approaches is the obvious loss of information. What is more, the results of clustering in these methods depend on the feature selection and problems with parametric modelling.

One of the most widely used methods of assessing similarity in the raw data approach is Dynamic Time Warping (DTW) (Berndt and Clifford [2]). Given two time series, $Q = q_1, q_{12}, ..., q_n$ and $R = r_1, r_2, ..., r_m$, DTW aligns them in such a way as to minimize their difference. The metric establishes an *n* by *m* cost matrix *C*, which contains the distances (Euclidean) between two points q_i and r_j. A warping path $W = w_1, w_2, ..., w_K$, where $\max(m,n) < K < m+n-1$, is formed by a set of matrix components, respecting three rules: boundary condition, monotonicity condition and step size condition. Eventually, the path that minimizes the warping cost is considered as DTW distance:

$$d_W(R,Q) = \min\left(\sqrt{\sum_{k=1}^{K} w_k}\right) \tag{1}$$

Dynamic programming is used for finding the path.

After determining the distance matrices, hierarchical or partitioning (crisp or fuzzy) clustering methods are used to find clusters. In order to evaluate an optimal number of clusters in the data, internal validity indices, such as the average silhouette width (Kaufman and Rousseeuw [9]), can be used. Adjusted Rand Index (ARI) (Hubert and Arabie [7]) can be next applied to compare the alternative classification results.

3. Data and empirical results

The data used in this study consist of monthly price indices from January 2001 to February 2014. All indices came from World Bank Commodity Price Data and are expressed in US dollars. Before the classification procedure, all price series are expressed as indices with their average values in 2007 equalling 1. The analysis is based on 54 series of variables, which are assigned to World Bank classes.

The whole sample period is divided into two sub-periods: 2001:1-2008:6 and 2009:1-2014:2, thus the classification is based on the pre-crisis and post-crisis periods[2]. The results are complemented by clustering series in the whole sample. The division is motivated by the disparate behaviour of commodity prices in these sub-periods.
DTW methods are used to classify time series. After obtaining dissimilarity metrics, Ward's, complete (hierarchical group of methods) and pam (division) methods are used to find cluster.

The results of clustering for the period 2001-2008 are presented in Figure 1[3], and they yield three main clusters of time series (the average silhouette width is the biggest for three clusters in Ward's method – see Table 1). The first cluster consists of 28 commodity prices including most energy commodities (their names in Fig. 1 begin with E), except for Gas US, metals (MM), (except for aluminium), and precious metals (PM). What is more, commodities belonging to the same category are close to each other, which means that their series paths are quite similar. The prices from the categories listed above are closest to one another, which means that their paths are most similar. The second cluster includes the prices of Gas US and Sugar EU, and it is hard to spot any connections between them. The last cluster consists of 24 commodity prices including most food, raw materials and beverages commodities.

[2] All the computations have been done using the R cran packages "dtw",,"TSclust", "cluster" and some others.
[3] The results of the complete method and pam method are available from the author upon request.

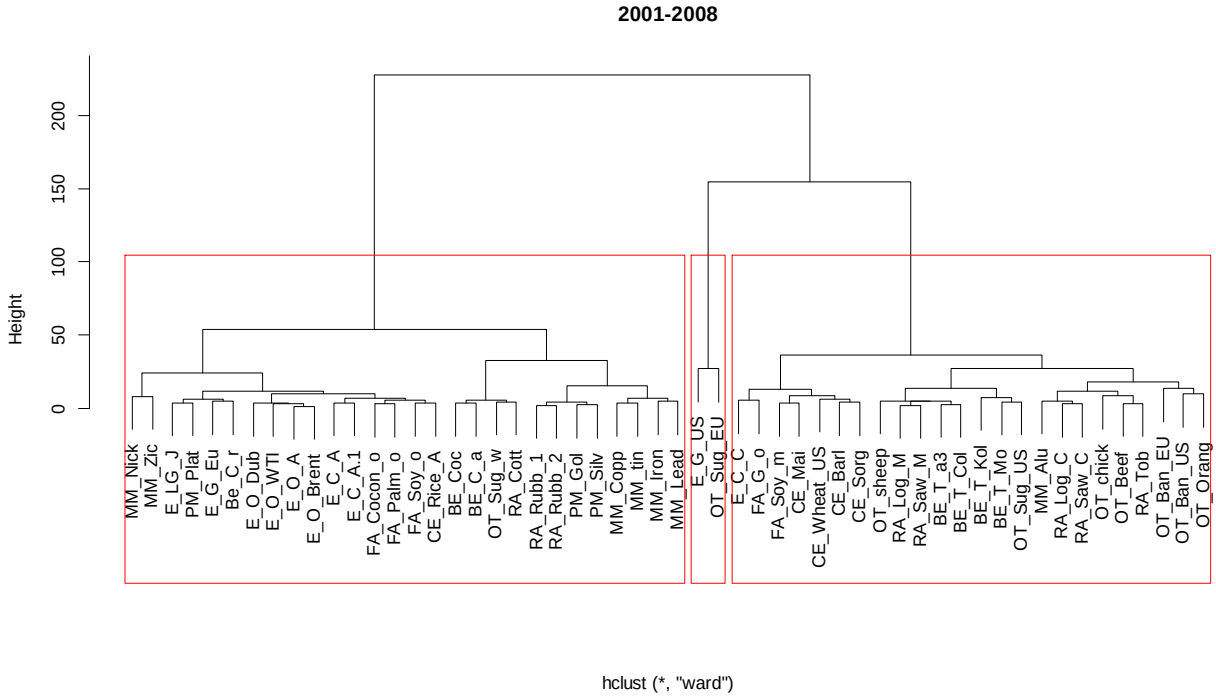


Figure 1 The classification results in the first sub-period

The results of clustering for the post-crisis period, with the assumption of Ward's method, are presented in Figure 2. Although in this case the average silhouette width suggests the division into 2 clusters, we have opted for three cluster and, as a result, energy, metal, and precious metals commodities are in different groups. There are 14 commodity prices in the first cluster, including oil prices (except for WTI Oil, which is in the second cluster), some food and raw material commodity prices. There are 19 elements in the second cluster, including most food and raw material prices, gold, and tin. There are 23 prices in the third cluster, including most precious metals, metals and minerals, coal prices and the remaining food commodities.

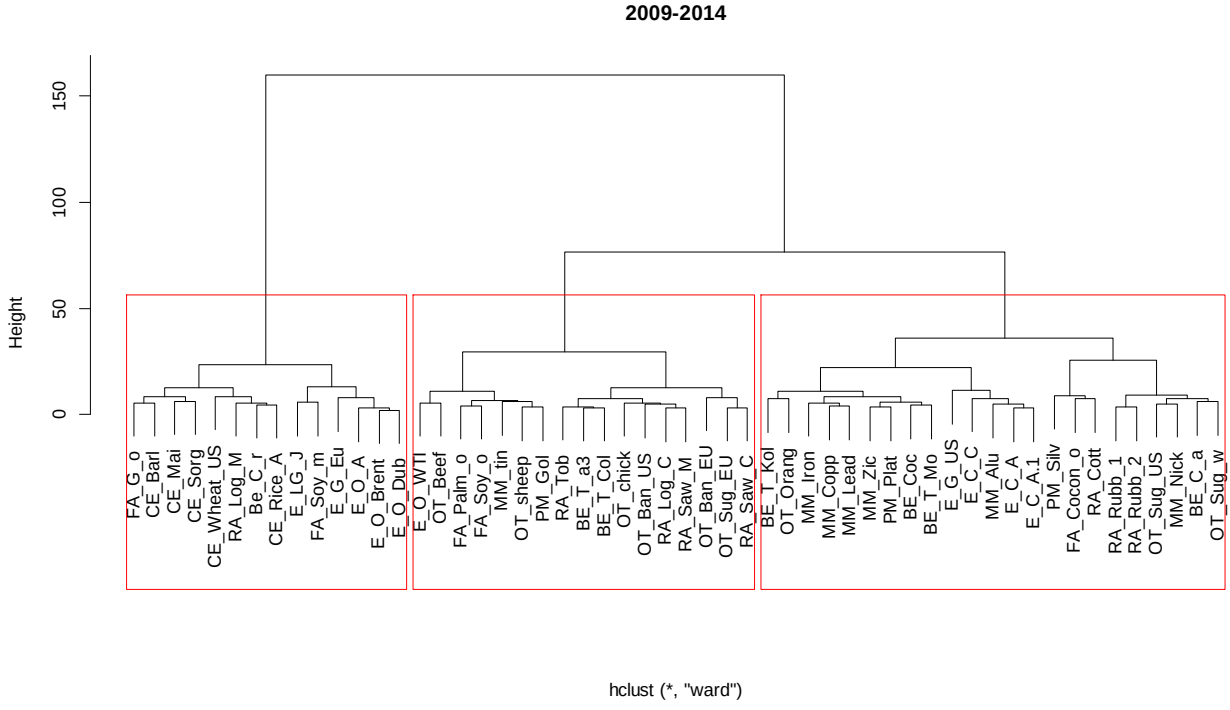


Figure 2 The classification results in the second sub-period

Finally, Figure 3 presents the results obtained for the whole sample. Here the average silhouette width suggests (see Table 1) the division into 4 groups (although the quality of division is rather poor). In the first cluster (17 elements) there are agricultural commodities (beverages, raw materials and other) and one industrial metal – aluminium. In the second cluster (18 elements) there are most industrial and precious metals, Australian Coal and some other agricultural

commodities. The last two clusters are quite close to each other. The third consists of US Gas and Sugar UE, while the fourth contains most energy commodities and some food, especially oils (palms, soya, groundnut).

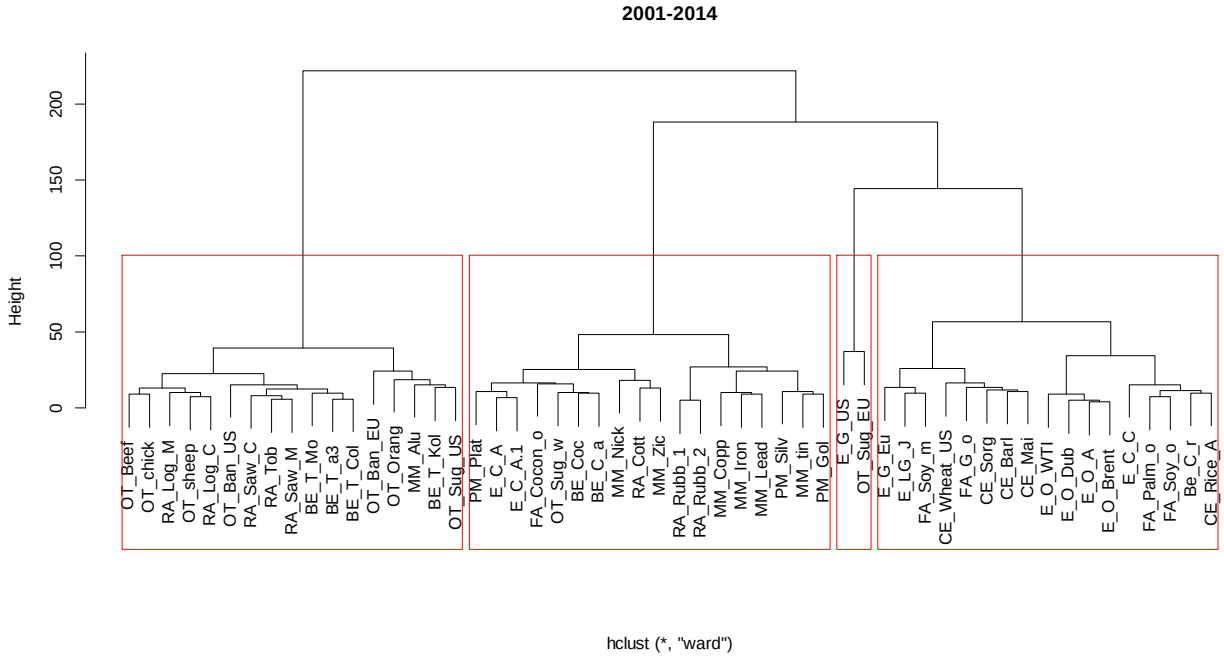


Figure 3 The classification results in the whole sample period

methods	**Ward's**			**complete**			**pam**		
period\nr cluster	2	3	4	2	3	4	2	3	4
2001-2014	0.235	0.287	0.322	0.605	0.265	0.322	0.274	0.196	0.225
2001-2008	0.374	0.427	0.301	0.746	0.423	0.274	0.382	0.429	0.301
2008-2014	0.371	0.257	0.221	0.378	0.237	0.221	0.338	0.245	0.214

Table 1 The average silhouette width for 2, 3 and 4 clusters in different time period

In order to compare the results of classifications, the adjusted rand index is computed (see Table 2). The level of agreement of different classifications and the comparison of clusters and categories of different commodities (listed in the World Bank indices – symbol WB in table 2) are measured. As there are six different categories of commodities in World Bank, the assumed division of the set of objects also consists of six clusters.

period	**2001-2014**			**2001-2008**			**2009-2014**		
	WB	ward	compl.	WB	ward	compl.	WB	ward	compl.
WB	1			1			1		
ward	0.100	1		0.139	1		0.077	1	
compl	0.123	0.543	1	0.163	0.626	1	0.074	0.467	1
pam	0.167	0.415	0.496	0.125	0.588	0.688	0.114	0.490	0.569

Table 2 Adjusted Rand Index for different classification methods

	Ward's		**complete**		**pam**	
			3 clusters			
periods	2001-2014	2001-2008	2001-2014	2001-2008	2001-2014	2001-2008
2001-2008	0.419	X	0.329	X	0.623	X
2009-2014	0.370	0.064	0.213	0.026	0.232	0.089
			6 clusters			
periods	2001-2014	2001-2008	2001-2014	2001-2008	2001-2014	2001-2008
2001-2008	0.730	X	0.406	X	0.318	X
2009-2014	0.238	0.156	0.331	0.130	0.192	0.084

Table 3 Adjusted Rand Index for different clustering methods and different periods

The results obtained indicate that WB commodity classifications differ greatly from clusters arising from statistical classification, which is clearly seen in low values of ARI for the first and the second (here the values are the highest) sub-periods as well as for the whole sample period. Thus, WB commodity indices do not determine co-movement of theirs elements. As far as various methods of obtaining clusters are concerned, the similarity of results are relatively high (from 0.467 obtained for pair *complete-ward* in second sub-period to pair *pam–complete* in the first sub-period). Again, higher values of the similarity measure are obtained in the first sub-period, which indicates that in this sub-period co-movement of indexes is more evident, and it is easily detected by different time series classification tools.

In order to compare the composition of clusters in different periods, ARI index is computed for 3 and 6 clusters and three classification methods. The results indicate (see Table 3) that the composition of clusters in pre-crisis and post-crisis periods are different. In the case of three clusters ARI varies from 0.023 for *complete* to 0.089 for *pam* method. In the case of six clusters the ARI is slightly bigger and reaches 0.156 for *Ward's* method. Relatively strong similarity of cluster composition in the pre-crisis sub-period and the whole period (ARI from 0.419 to 0.73 for 6 clusters) results from the fact that co - movement of all commodity prices in the pre-crisis period is stronger and more evident.

4. Conclusion

Dynamic time warping is used in the study to classify commodity price data in the pre-crisis and post-crisis periods. The results obtained reveal that co-movement of commodity prices is more evident in the pre-crisis period, when the clusters are more homogeneous and consist of commodities from the same category (e.g. precious metals or energy commodities are located in the same cluster). Clusters obtained for the post-crisis period are less homogeneous. The internal classification measure demonstrates that the best division is obtained if only two or three clusters are considered in every period. Clusters obtained for the whole period sample indicate that there are only two patterns of behaviour of prices in the periods analysed (stronger in the first one). Comparing commodity categories with the results of clustering indicates that commodities which belong to one category do not always behave in the same way. It is especially evident in the second period, when certain energy commodities, metals or precious metal belong to different clusters.

The results obtained might be of great importance to investors, as they demonstrate that at present co-movement of commodity prices is not as evident as it used to be. What is more, a well-diversified portfolio can consist of commodities from the same classes.

Acknowledgements

Supported by the grant No. 2012/07/B/HS4/00700 of the Polish National Science Centre.

References

[1] Akram, F.Q.: Commodity prices, interest rates and the dollar. *Energy Economics* **31** (2009), 838-851.
[2] Berndt D., and Clifford. J., Using dynamic time warping to find patterns in time series. *KDDworkshop*, Vol. 10, **16** (1994), 359-370.
[3] Byrne, J.P., Giorgio Fazio, and G., Fiess N.: Primary commodity prices: Co-movements, common factors and fundamentals, *Journal of Development Economics* **101** (2013), 16-26.
[4] Frankel, J.A.: The effect of monetary policy on real commodity prices. In: *AssetPrices and Monetary Policy* (John Campbell Y.J., eds.), NBER, University of Chicago, Chicago. 2008. NBER Working Paper 12713.
[5] Gilbert, C.L.; Speculative Influences on Commodity Futures Prices, 2006-2008, *Working Paper*, Department of Economics, University of Trento, Trento, 2009.
[6] Gohin, A., and Chantret, F.,: The long-run impact of energy prices on world agricultural markets: the role of macro-economic linkages. *Energy Policy* **38** (2010), 333–339.
[7] Hubert, L., and Arabie, P.: Comparing partitions. *Journal of Classification* **2** (1) (1985), 193- 218.
[8] Irwin, S.H., and D.R. Sanders.: Index Funds, Financialization, and Commodity Futures Markets, *Applied Economic Perspectives and Policy* **33** (2011):1-31.
[9] Kaufman, L., and Rousseeuw, P.J.: *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley & Sons, New York, 1990.
[10] Krugman, P., Commodity Prices, NYTimes, March 19, 2008.
[11] Liao, W.,T.: Clustering of time series data—a survey, *Pattern Recognition* **38** (2005), 1857 – 1874.
[12] Natanelov, V., Mohammad, J., Alam, M.J., McKenzie, A.M., and Huylenbroeck G.V.: Is there co-movement of agricultural commodities futures prices and crude oil? *Energy Policy* **39** (2011), 4971–4984.
[13] Nazlioglu, S, and Soytas, U.: Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis, *Energy Economics* **34** (2012),1098–1104.
[14] Papież, M., and Śmiech, S.: The Analysis of Relations Between Primary Fuel Prices on the European Market in the Period 2001-2011, *Rynek Energii* **5** (2011), 139-144**.**

[15] Phillips, P. C. B., and J. Yu.: Dating the Timeline of Financial Bubbles during the Subprime Crisis. *Cowles Foundation Discussion Paper No. 1770*, Yale University, Yale, 2010.

[16] Svensson, L.E.O.: The effect of monetary policy on real commodity prices: Comment. In Asset Prices and Monetary Policy. Ed. John Y. Campbell, NBER, University of Chicago, Chicago. 2008. And NBER Working Paper 12713.

Methods of solving missing data issues in credit risk scoring and comparison of its effectiveness

Nataliya Soldatyuk[1], Stanislav Sopko[2]

Abstract. The model assessing and predicting the credit risk is very sensitive to the input data. Data about consumers are often incomplete, especially because application forms are very different in banks, it makes data sample inconsistent. This is causing difficulties in building a model based on this data. Usually to solve these issues incomplete records are left out from the observation dataset. As a result of this approach we lose important information and the model does not cover entire spectrum of consumers. It does reduce the quality and the predictive power of the model. There are several methods solving missing data issues without losing part of data: substitution approaches, the maximum likelihood method, the multiple imputations and other. We have applied these methods on live data sample. The work's aim is to evaluate and compare effectiveness of these techniques.

Keywords: credit scoring, missing data, EM algorithm, maximum likelihood, multiple imputation.

JEL classification: C44
AMS classification: 90C15

1 Introduction

In order to reduce the credit risk in banking sector Basel II recommends using models based on internal and external data to estimate credit risk components such as probability of default (PD), loss given default and exposure at default. Central banks in many countries have released a directive to make a central credit bureau with data from all banks in a country. Credit scoring models are widely used in banking practice. First of all predictive power of the model depends on the quality of the data. The central credit bureau receives information from various banks and the problems with data completeness are exposed at this stage. Usually some records are incomplete or contain inadequate historical data. There are a lot of reasons for that: application forms are very different in banks, business rules allow clients leave some fields blank, mechanical or system errors, fake data. Such records can not be deleted because it is particularly important to have a full range of information to build the model. Taken into account that we want use the available information as efficiently as possible, we decided to replace the missing values with their estimation using mathematical methods.

The goal of this article is to find the most appropriate method for substitution of missing values. For this purpose we have focused on a maximum likelihood and multiple imputation methods, which we have applied on a data sample with missing values. The result of this analysis was two data samples with substituted data, which we have compared with the original data sample with no missing data.

2 Missing data

There are three classifications according to [2],[10] depending on the nature of missing data. Lets consider the following notation: A - data matrix $(n \times K)$ built on a sample of n observations, y_i univariate outcome and x_i is a $p \times 1$ vector of covariates corresponding to y_i. Assume that y_i is completely populated while some components of x_i $(i = 1, ..., n)$ are missing for observation i. Denote by R_i the vector of response indicators (ie, $R_{ij} = 1$ if x_{ij} is populated, 0 if this element is not populated).

[1]University of Economics, Department of Econometrics, W. Churchill Sq. 4, Prague, Czech Republic, xsoln900@vse.cz
[2]University of Economics, Department of Econometrics, W. Churchill Sq. 4, Prague, Czech Republic, stanislav.sopko@vse.cz

Missing data are missing completely at random (MCAR) if failure to observe a value does not depend on any values of variables, either observed or missing. The missing values of x_i are MCAR if the probability of observing x_i is independent of value y_i and of the values of components of x_i that are observed or would have been observed. In this case, vector R_i is a random vector of zeros and ones. For example, if we don't have data about age of some customer, the fact that age is missing does not depend on age of this person or any other parameters.

Missing at random (MAR) data present a situation when the probability of observing data is independent of values of these data, however conditional probability of missingness may depend on any observed data. For example, probability of observing components of x_i is independent of the values of x_i that would have been observed, but this probability is not necessarily independent of y_i and observed components of x_i. If data are MCAR, then they are MAR.

If the failure to observe a value depends on the value that would have been observed, the missing data said to be missing not at random. There is dependence between values of components of vector R_i and missing values of x_i.

We assume that our data, which described in this paper are MAR.

3 Data

For our research we have used 4100 observations of credit applications. Initially the data did not contain missing values. We have removed a part of values randomly and thus received a sample which was used for the analysis. It allows us to compare substituted data with initial values and determine efficiency of the method. Data sample contains 10 exogenous variables (8 of them are categorical and 2 are continuous) and one class attribute. According to [1] we assume that missing data affect not all but only some variables. These variables have been chosen randomly: Al, A2, A3, A6, A7 and A8 contain missing values (55.00% of exogenous attributes). We have set the percentage of observations with missing data to less than 10% based on the recommendations described in [1]. As mentioned in [8], the key point for the practice is the fact that the imputation model contains at least the same amount of information as the analysis model. The sample must include at least $p(p+3)/2$ observations, where p is the number of exogenous variables ($p = 11$). In our case, number of observations in the sample exceeds this value more than 50 times. The input data are described in detail in the Table 1.

Var	Nature,range	# Missing,(%)	Mean(Median)
A1	Categorical, $\{1, 2\}$	62,(1,508)	1
A2	Categorical, $\{1, 2, 3, 4, 5\}$	57 (1,387)	2
A3	Categorical, $\{0, 1, 2, 3, 4, 5\}$	42 (1,022)	0
A4	Categorical, $\{0 - 23\}$	0	8
A5	Categorical, $\{0, 1, 2, 3, 4\}$,class attribute	0	1
A6	Categorical, $\{0 - 9\}$	25 (0,608)	1
A7	Categorical, $\{1 - 8\}$	13 (0,315)	5
A8	Categorical, $\{0 - 58\}$	9 (0,22)	20
A9	Continuous, $(0, +\infty)$	0	4525981
A10	Categorical, $\{2 - 71\}$	0	14
A11	Continuous $(0, +\infty)$	0	2580185

Table 1 Input data statistics

4 Methods of solving missing data issues

4.1 Maximum Likelihood method

Maximum Likelihood method suggests the values for the missing data based on estimation of parameters of the joint distribution,[2]. Detailed ML method described in [2] and [3]. Let's assume that our missing values are missing at random. Then we have a sample of n independent and identically distributed

observations. Maximum Likelihood (ML) method firstly defines the joint density function for all records. In this step we are making estimation of the main statistics of the distribution (mean and the covariance matrix) by calculation of the sample log-likelihood function. Let's assume, that our data have multivariate normal distribution. The log likelihood for the sample is the sum of individual log-likelihood values:

$$\log L = \sum_{i=1}^{n} \log l_i = \sum_{i=1}^{n} \log\{\frac{1}{(2\pi)^{K/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(z_i-\beta)^T\Sigma^{-1}(z_i-\beta)}\}, \tag{1}$$

where $\log L$ is the sample log likelihood function, $\log l_i$ is an individual log likelihood function for observation i, K is the number of variables, z_i is the score vector for observation i ($z_i = (x_i, y_i)$), β is the population mean and Σ is the covariance matrix. The estimation procedure is an iterative process that repeatedly calculates the sample log likelihood every time substituting different values of β and Σ to maximize the likelihood. By this it finds the distribution, which might most likely produce the sample. There are specific methods for finding optimal estimation of the joint distribution such as the expectation-maximization algorithm ([1],[4]).

Then, using estimates of statistics of distribution, the method calculates the log likelihood value for each observation to find the best substitution for missing data. With an assumption of a multivariate normal distribution for the population, case i log-likelihood is

$$\log l_i = -\frac{K}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(z_i-\beta)^T\Sigma^{-1}(z_i-\beta) \tag{2}$$

where $(z_i-\beta_i)^T{\Sigma_i}^{-1}(z_i-\beta_i)$ is the Mahalanobis distance value, which calculates the standardised distance between the data points of i observation and the center of the multivariate normal distribution. If some data for observations i are missing, then we need to take into account fact that for this observation number of populated variables is less then K.

The log-likelihood in case of missing data with regard to changes of number of variables for the specific case is:

$$\log l_i = -\frac{k_i}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(z_i-\beta_i)^T{\Sigma_i}^{-1}(z_i-\beta_i) \tag{3}$$

The above stated equation defines the relative probability that an observation i comes from a multivariate normal distribution with specific mean and covariance matrix. Here k_i ($k_i \leq K$) is the number of populated variables for this observation, β_i and Σ_i are statistics of distribution regard to changes of number of variables. The main aim for the estimation is to find the combination of values which gives the highest log-likelihood and so fits to the data best.

As we mentioned, one of the techniques to find the joint distribution is the optimization EM algorithm. The algorithm is a two-step iterative process. Detailed technique of EM algorithm described in [4]. The first step (E-step) uses initial estimation of the mean vector and covariance matrix to build regression equations that predict the incomplete variables, and the second step (M-step) afterwards produces updated estimates of β and Σ. This two-step process is repeated until β and Σ no longer change and converged on the maximum likelihood estimates. EM algorithms mission is not to determine values of missing data, but to find the optimal estimation of β and Σ. Let's $\beta^{(t)}$ be the estimation of β at the current iteration. For a given observation i, x_i could be divided into two components $x_i = (x_{obs,i}, x_{mis,i})$, where $(x_{obs,i})$ presenting observed components and $(x_{mis,i})$ denoting missing components of x_i.

The E-step calculates the expected complete sample log-likelihood ratio. The E-step for all observations can be written as

$$Q(\beta, \beta^{(t)}) = \sum_{i=1}^{n} E(\log l_i(\beta)|x_{obs,i}, y_i, \beta^{(t)}) = \sum_{i=1}^{n}\sum_{x_{mis,i}} p(x_{mis,i}|x_{obs,i}, y_i, \beta^{(t)}) \log l_i(\beta), \tag{4}$$

where $p(x_{mis,i}|x_{obs,i}, y_i, \beta^{(t)})$ is the conditional distribution of the missing covariates given the observed data and current estimate of β; $\log l_i(\beta)$ is log likelihood function calculated using β. Sum is taken over all possible values of $x_{mis,i}$.

The M-step consist of maximizing $Q(\beta, \beta^{(t)})$ in β for fixed $\beta^{(t)}$

$$\beta^{(t+1)} = arg\max_{\beta} Q(\beta, \beta^{(t)}) \tag{5}$$

We have applied the ML method on our dataset for all variables of both types (categorical and continuous). Estimation for values of missing data received by the ML method is proven to be effective. However, continuous variables A9 and A11 have caused a situation, that every record has been observed as unique combination. As a result we have obtained very low values of the log-likelihood function (mean value of the likelihood function is -49.9016). This is not a critical obstacle to the analysis, nevertheless we decided to discretize these variables by the following rule:

$$\begin{cases} A9' = 0, & A9 = 0, \\ A9' = 1, & A9 \ \epsilon \ (0, 5000), \\ A9' = 2, & A9 \ \epsilon \ [5000, 10000), \\ A9' = 3, & A9 \ \epsilon \ [10000, 20000), \\ A9' = 4, & A9 \ \epsilon \ [20000, +\infty). \end{cases}$$

Using the same logic but another corresponding intervals, we categorized variable A11.

For estimation of the joint conditional distribution we used EM algorithm. Resulting values of mean and covariance matrix have a very small deviation from corresponding values of the complete data sample. We have been expecting this deviation to be small considering that missing data affect only approximately 6% of all observations. Table 2 represents the deviation in percents of the values estimated by EM algorithm from the corresponding values of the complete data sample. For better comparison we have also added a column with a number of missing data in percents for each variable.

	% of missing data	$(\beta - \beta_{est}), \%$	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
A1	1.508	0.014	0.097										
A2	1.387	-0.059	-1.868	-0.479									
A3	1.022	3.139	-0.071	2.346	3.290								
A4	0	0	-1.015	4.828	2.652	0.024							
A5	0	0	-11.135	-23.175	1.732	0.024	0.024						
A6	0.608	0.010	-8.353	-3.469	-0.175	-0.976	0.049	-0.130					
A7	0.315	0.019	-3.556	-6.894	-0.934	0.241	0.147	-0.731	0.041				
A8	0.22	0.016	-1.527	-0.784	-1.832	0.880	0.156	-0.808	-0.241	0.010			
A9	0	0	-0.875	-1.071	4.366	0.024	0.024	-0.551	0.823	2.017	0.024		
A10	0	0	-3.494	-13.195	2.640	0.024	0.024	-0.321	0.037	1.371	0.024	0.024	
A11	0	0	-1.124	-12.056	65.765	0.024	0.024	-0.396	0.259	1.833	0.024	0.024	0.024

Table 2 Input data statistics

Results show good performance of EM algorithm. The estimates of mean values for variables A1, A6, A7, A8 biased from the real values less than 0.02%. However, for the variable A3 bias of estimation from the real value reached 3.14% whereas this variable contains only 1.02% of missing values. The reason for this is that in reality the distribution of the variable A3 deviates from the normal distribution, while EM algorithm's estimations are made with an assumption of normal distribution for all variables.

We have used the estimates obtained by EM algorithm as initial statistics of distribution for subsequent phase. We calculate the log likelihood function for each observation by formula (3). Substituting all possible values of the missing parameters, we have chosen those values which maximize the sample log likelihood. As a result we have obtained the sample log likelihood $\log L = -83956.4$ (log likelihood for initial complete data sample $\log L = -84090.5$). Resulting estimation of ML method is compared with the statistics of the initial complete sample. It is described in the Table 3.

The results described in the Table 3 show the high efficiency of ML method. The resulting data sample certainly can be used for subsequent building of a scoring model. Taking into account that our variables are categorical, we can evaluate the accuracy of the method by count of the number of exact matches. From the total number of 208 record with missing data we have obtain 125 exact matches (60%) and 47 of them had deviation of 1 point from the real value.

From the practical point of view it is very important to substitute missing data by adequate values. Random combinations of parameters values can give us a case, which will never appear in life (for example, estimation, that 18 old individual have high level education and 3 children). In this sense, ML suggests

	Complete data					Maximum Likelihood				
Variable	β	Std. Dev	σ^2	95% Confidence Limits		β_{est}	Std. Dev	${\sigma^2}_{est}$	95% Confidence Limits	
A1	1.337	0.473	0.224	1.323	1.352	1.333	0.471	0.222	1.318	1.347
A2	1.681	0.615	0.378	1.662	1.700	1.685	0.614	0.376	1.666	1.704
A3	0.176	0.737	0.542	0.154	0.199	0.169	0.720	0.518	0.147	0.191
A6	1.395	0.775	0.600	1.371	1.419	1.395	0.774	0.598	1.372	1.419
A7	4.842	3.337	11.137	4.740	4.944	4.841	3.333	11.112	4.739	4.943
A8	19.944	9.351	87.433	19.658	20.230	19.942	9.342	87.269	19.656	20.228

Table 3 Input data statistics

potential substitution for missing values similar to existing combinations in the sample. Although such substitutions do not carry a lot of new information for further sample analysis, but it certainly increases statistical power of the sample.

4.2 Multiple imputation method

Multiple imputation (MI) method replaces each missing value by $m > 1$ simulated values, where m could be close as small as 5-10,[10]. As a result, there are m sets of imputations which are plausible versions of the complete data set. Every data set is analysed using complete-case methods. Results are combined to produce only one optimal solution. Rubin ([10]) developed multiple imputation in the Bayesian framework. The multiple imputation process consists of 3 different phases: imputation phase, analysis phase, and pooling phase. Based on an initial estimate of β and Σ the imputation phase builds a set of regression equations. These equations are intended to predict the incomplete statistics from the observed variables. The imputation equation is ${X_i}^* = [\hat{\beta}_0 + \hat{\beta}_1(Y_i)] + z_i$, where z_i is a random residual from a normal distribution with mean of zero and variance equal to the residual variance from the regression. The goal of the pooling step is to produce alternate estimates of β and Σ. Next, the algorithm generates a new set of parameter values by adding a random residual to each element in $\hat{\beta}$ and $\hat{\Sigma}$.

Using MI method, we have generated 5 data sets of imputations with a possible values of missing data. We decided to operate with 5 data sets according to [7] and [10], where it is described, that for moderate amounts of missing data, it's possible to get over 90% efficiency with just 5 data sets. In the Table 4 is presented an optimal estimation of β and standard errors obtained by multiple imputation method. These values are compared with β and standard errors for the case of initial complete data.

	Complete data					Multiple Imputation				
Variable	β	Std. Dev	σ^2	95% Confidence Limits		β_{est}	Std. Dev	${\sigma^2}_{est}$	95% Confidence Limits	
A1	1.337	0.473	0.224	1.323	1.352	1.338	0.474	0.224	1.323	1.352
A2	1.681	0.615	0.378	1.662	1.700	1.682	0.619	0.383	1.663	1.701
A3	0.176	0.737	0.542	0.154	0.199	0.170	0.727	0.528	0.148	0.192
A6	1.395	0.775	0.600	1.371	1.419	1.396	0.776	0.602	1.372	1.420
A7	4.842	3.337	11.137	4.740	4.944	4.843	3.336	11.131	4.741	4.945
A8	19.944	9.351	87.433	19.658	20.230	19.942	9.348	87.393	19.656	20.228

Table 4 Input data statistics

Similarly to the case of ML method the results show the high efficiency of the MI method. An estimation of the mean have lower deviation from the real values than in the case of ML method. Nevertheless, the number of exact matches by MI shows worse result then ML: 80 exact match (39%) from a total number of 208 missing data.

In case if it is possible to select sufficient number of observations without missing data, it's not necessary to use the analyzed substitutional techniques. Otherwise if the data are insufficient and observations with missing data have to be included into the development sample, those methods can be very helpful. Another interesting usage of these methods is an identification of certain types of individuals. One of the problems in lending, which brings big losses to banks is a fraud. Credit applications completed by applicants carry a lot of information, and fields with missing values bring important information as well as fields with existing values. If fraudsters deliberately are not filling some fields, an investigation

and restoration of the missing values might increase the probability of fraud detection at the time of application.

5 Conclusions

In the study described in this article, we have applied maximum likelihood and multiple imputation methods for solving missing data issue. As an outcome of this analysis we have received two data samples with substituted data, which we have compared with the original data sample with no missing data. Both techniques are good approaches and both of them give effective results. Taking into account all specifics of the data, there might be identified benefits and deficiencies of these methods. For our data, ML is more efficient than MI because it generates higher number of matches with real values. Besides this, ML method is easier to use because for a given set of data ML method produces same results every time it is applied. Since MI uses random draws, the results contain some level of indeterminacy. When the method is applied to the same data set it produces different standard errors, parameter estimates, and test statistics every time. According to [9] to get fully efficiency, you would have to produce and analyze an infinite number of data sets, which is not possible.

We are going to continue our research and the next step will be building of scoring model based on the sample with replaced values. After that we will compare predictive power of that model with model built without replaced values. Also we will analyze some other interesting methods for solving missing data issue, such as fully Bayesian approach and weighting method.

Acknowledgements

This paper supported by the grant No. 58/2014, IG403024 by the Internal Grant Agency of the University of Economics Prague.

References

[1] Florez-Lopez,R. : *Effects of missing data in credit risk scoring. A comparative analysis of methods to achieve robustness in the absence of sufficient data. Journal of the Operational Research Society*, **61.** (2010), pp.486-501.

[2] Ibrahim,J., Chen, M., Lipsitz, S., and Herring, A., : *Missing data methods for generalized linear models: A comparative review. J Am Stat Assoc*, **100(469)** (2005), pp.332-346.

[3] Enders, C.: *Applied Missing Data Analysis* The Guilford Press, New York, (2010) ISBN: 978-1-60623-639-0.

[4] Witten, I., Frank, E. : *Data Mining. Practical Machine Learning Tools and Techniques, Second Edition.* Elsevier Inc. (2005), ISBN: 0-12-088407-0.

[5] Mays, E.: *Handbook of Credit Scoring.* Amacom, New York. (2001), ISBN: 0-8144-0619-X.

[6] Verstraeten,G., Van den Poel,D.: *The Impact of Sample Bias on Consumer Credit Scoring Performance and Profitability. The Journal of the Operational Research Society*, **56.** (2005), pp. 981-992.

[7] Tufis, C.: *Multiple imputation as a solution to the missing data problem in social science. Calitatea Vietii*, XIX, **12**, (2008), p. 199-212.

[8] King, G., Honaker, J., Joseph, A. and Scheve, K. : *Analyzing incomplete political science data: An alternative algorithm for multiple imputation. Am Polit Sei Rev*, **95(1)**: (2001), p.49-69.

[9] Allison, P.: *Handling missing data by maximum likelihood. Statistical Horizons, Haverford, USA.* **312**: 2012.

[10] Rubin, D.: *Multiple Imputation for Nonresponse in Surveys. SWiley and Sons: New York*, (1987), ISBN 0-471-08705-X.

Using cluster analysis techniques based on K-means and Kohonen clustering methods in credit scoring

Stanislav Sopko[1], Nataliya Soldatyuk[2]

Abstract. Methods using cluster analysis are very useful in data analysis, especially when the studied sample data are big. The main focus of this paper is studying the use of cluster analysis as a part of a predictive algorithm in the area of credit scoring. This methodology is applied subsequently: first by determining to which cluster the client belongs, second by calculating a specific credit risk scorecard for each cluster. It is expected that this approach provides better results than using a single scorecard for all the clients in the sample. The aim of this article is to build a useful model for banking field by dividing the groups of borrowers into clusters in order to obtain a profile of the customers (debtors and good payers). We assume that a class (cluster) is appropriate if it contains members that have a high degree of similarity measured by a variance in the cluster. Techniques used in this paper are primarily based on K-means algorithm and Kohonen clustering method. K-means algorithms as an iterative technique reaching the optimum clustering are used for assignation of a set of n observations into k clusters. Results, acquired after application of K-means, are compared with results from SOM (self-organizing map) and Kohonen clustering method. The basis of Kohonen cluster is a neural network but it has a several similarities with K-means and comparison of these two approaches should result in interesting conclusions.

Keywords: cluster analysis, K-means, Kohonen clustering method, SOM, big data, credit scoring

JEL Classification: C44
AMS Classification: 62H30

1 Introduction

This paper deals with big data and possibilities of using of clustering methods in this area. Basic principles of clustering, K-means algorithm, Kohonen clustering method and neural networks are described in the first part. In the second part there are computational experiments with proposed solution on data sample. The main focus of this paper is to create a background for further analysis and research in the area of clustering of big data and especially using K-Means and Kohonen algorithms.

2 Theoretical basis

The theoretical basis of this paper is primarily based on the theory and findings from [1], [3], [6], [10], [12] and [13]. In addition, we also used sources [2], [4], [5], [7], [8], [9], [11] to complement the primary theoretical basis.

2.1 Cluster analysis

Cluster analysis comprises of a range of methods for classifying multivariate data into subgroups. By organizing multivariate data into such subgroups, clustering can help reveal the characteristics of any structure or patterns present. These techniques have proven useful in a wide range of areas such as medicine, psychology, market research and bioinformatics.

In other words it is statistical classification technique in which cases, data, or objects (events, people, things, etc.) are sub-divided into groups (clusters) such that the items in a cluster are very similar (but not identical) to one another and very different from the items in other clusters. It is a discovery tool that reveals associations, patterns, relationships, and structures in masses of data.

[1] University of Economics, Department of Econometrics, nám. W. Churchilla 4, Prague, Czech Republic, stanislav.sopko@vse.cz.
[2] University of Economics, Department of Econometrics, nám. W. Churchilla 4, Prague, Czech Republic, xsoln900@vse.cz.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. There are a lot of algorithms that can divide the monitored data into relevant groups. We choose from the specific reasons K-means algorithm and Kohonen clustering method based on artificial intelligence. These specific reasons are justified by the fact that these techniques are used in a many professional clustering software used in the current time.

K-means algorithm

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define *k* centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate *k* new centroids as centers of the clusters resulting from the previous step. After we have these *k* new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the *k* centroids change their location step by step until no more changes are done. In other words centroids do not move any more.

Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_i^{(j)} - c_j||^2$$

where $||x_i^{(j)} - c_j||^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centres.

The algorithm is composed of the following steps:

1. Place *K* points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the *K* centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

K-means is a simple algorithm that has been adapted to many problem domains.

Kohonen clustering algorithm and self-organizing maps

Kohonen Self Organizing Feature Maps, or SOMs provide a way of representing multidimensional data in much lower dimensional spaces - usually one or two dimensions. This process, of reducing the dimensionality of vectors, is essentially a data compression technique known as vector quantization. In addition, the Kohonen technique creates a network that stores information in such a way that any topological relationships within the training set are maintained. One of the most interesting aspects of SOMs is that they learn to classify data without any external supervision whatsoever. It consists of neurons or map units, each having a location in a continuous multi-dimensional measurement space as well as in a discrete two dimensional data collection is repeatedly presented to the SOM until a topology preserving mapping from the multi-dimensional measurement space into the two dimensional output space is obtained. This dimensionality reduction property of the SOM makes it especially suitable for data visualization.

There are „*m*" cluster units, arranged in a one or two dimensional array and the input signals are *n*-tuples. The weight vector for a cluster unit is the exemplar of the input patterns associated with that cluster. In self-organizing process, the cluster unit whose weight vector matches the input pattern closely is selected as the winner. The winning and the neighboring units update their weights.

$$c = \min_i ||X - m_i||$$

In which m_i is the location of the *i*-th map unit in the measurement space and *c* is the index of the winner map unit in the output grid of SOM. After the winner search, the locations of the map units in the measurement space are updated according to the rule:

$$m_i(t+1) = m_i(t) + \alpha(t) h_{ci}(t)[x(t) - m_i(t)] \quad (1)$$

$\alpha(t)$ is from interval (0,1) and represents a learning factor, $h_{ci}(t)$ is the Gaussian neighborhood defined as

$$h_{ci}(t) = \alpha(t).\exp(-\frac{||r_i - r_c||^2}{2\sigma(t)^2})$$

The neighborhood function includes the learning rate function $\alpha(t)$ which is a decreasing function of time and the function that dictates the form of the neighborhood function. The form of the latter function also determines the rate of change around the winner unit. A variety of neighborhood functions can be used. We can constrain the neighborhood function to be non-increasing around the winner unit m_c. Thus, the neighborhood function can also be constant around the winner unit. So r_c is the location of the winner unit and r_i is the location of the i-th map unit in the discrete output grid of SOM. $(t)$and $\sigma(t)$ are monotonically decreasing functions of time t.

3 Proposed algorithm, experimental analysis

K-Means fails to give optimum result when it comes to clustering high dimensional data set because its complexity tends to make things more complicated when more number of dimensions is added. In Data Mining this problem is known as "Curse of High Dimensionality". The main complication is that the exact number of cluster is not known prior to clustering. We propose a modified K-means algorithm as an appropriate combination of Kohonen SOM and standard K-means.

3.1 Modified K-means algorithm

The proposed algorithm consists of two parts. In the first one we find an exact number of clusters using Kohonen SOM. In the second part we take known number of clusters K and execute standard K-means. We basically apply KSOM on the dataset for reducing the dimension, keeping intact the topological structure of the data. KSOM not only reduces the dimension but it also gives an exact number of clusters. The initial centroids are randomly selected in each cluster and the reduced data are grouped into cluster using K-means algorithm. The exact steps of the two mentioned parts of proposed algorithm are following:

1. part
1. step: Get a set of data of n items with m attributes
2. step: Each node's weights are initialized
3. step: A vector is chosen at random from the set of training data and presented to the lattice
4. step: Every node is examined to calculate which one's weights are most like the input vector. The winning node is commonly known as the best matching unit (BMU). Euclidean distance is used to find similarity.
5. step: The radius of the neighborhood of the BMU is now calculated
6. step: Each node's (found in the previous step) weights are adjusted to make them more like the input vector. The closer a node is to the BMU, the more its weights get altered (according to (1))
7. step: Repeat 3. step for N iterations

2. part
1. step: Get set of n data points with m attributes, number of desired clusters K
2. step: Randomly selected K initial centroids for each cluster
3. step: Assign each point a_i to the cluster which has the closest centroid. Calculate the new mean for each cluster until the convergence criteria is met.
4. step: Repeat 3. step

The proposed algorithm will be implemented in Matlab, which also has its own tools for modeling of neural networks and artificial intelligence. The results will be compared with results obtained with non-modified algorithms.

3.2 Data sets

To test the performance of proposed algorithm, Australian credit approval data set from UCI machine learning repository is used. Data set contains 14 attributes, 6 numerical and 8 categorical. This file concerns credit card applications. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. The data are interesting because there is a good mix of attributes – continuous, nominal with small numbers of values, and nominal with larger number of values. There are also a few missing values. But first, we choose a data set of 1 000 members with only 2 attributes for better understanding our proposed algorithm and visualization in 2 dimensional space (Matlab's simplecluster_dataset). There are methods that can reduce dimensionality of many problems, so that proposed data set can represent real problem in the area of credit scoring. After that we applied proposed algorithm to Australian credit approval data.

3.3 Experimental analysis

The algorithm is implemented in the environment of Matlab. First, we implement Kohonen SOM a then standard K-means. As we mentioned, we use as input 2 data groups - simple data sample for demonstration of functionality of proposed algorithm and real data of Australian credit approval.

Simple cluster dataset

We use data, which Matlab offers, as an input for self-organizing map. Then we train constructed net by 200 iterations and monitor changes, weights and adaptation of net. The final visualization of Kohonen SOM is captured in Figure 1, Figure 2 and Figure 3.

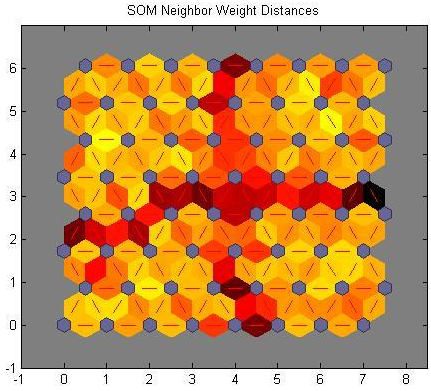


Figure 1 SOM weight distances – simple cluster dataset

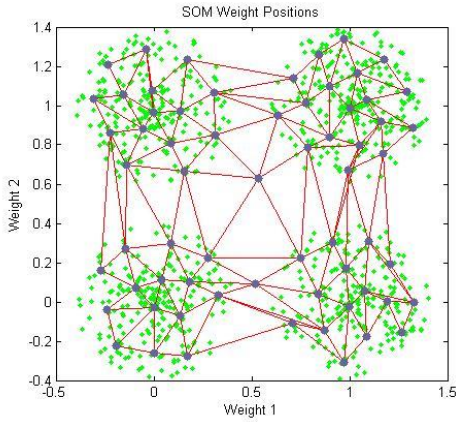


Figure 2 SOM weight positions – simple cluster dataset

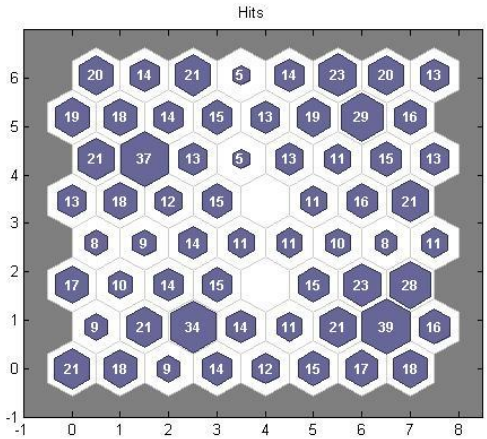


Figure 3 Hits of SOM – simple cluster dataset

Kohonen SOM and its visualization help us to clearly define optimal number of clusters. The optimal number is 4 and we use this parameter as input for the phase 2 and K-means algorithm. The result is captured in Figure 4.

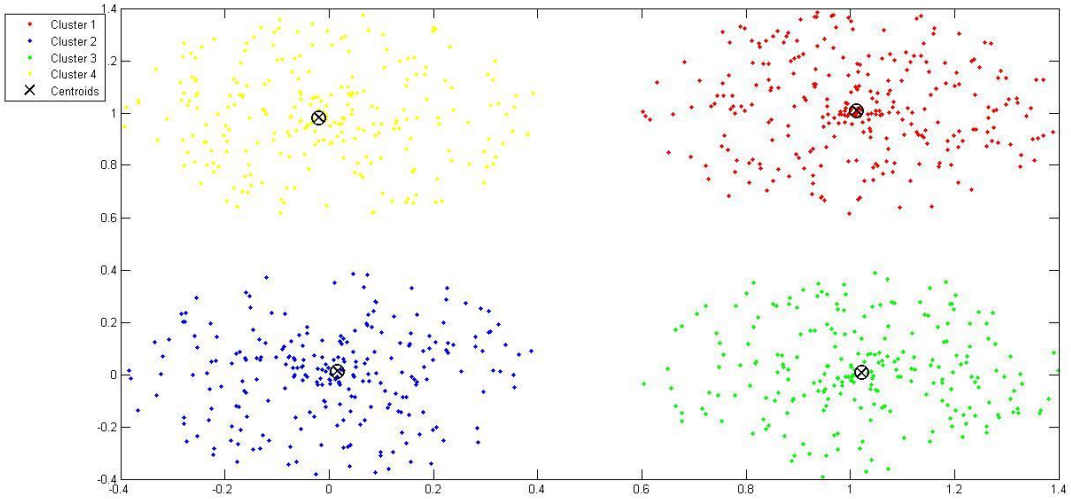


Figure 4 Results after K-means application (with 4 clusters)

Australian credit approval data

We use very similar procedure in this case. Data sample contains 690 members with 15 attributes (some of values are missing by design). SOM graphic outputs are captured in Figure 5.

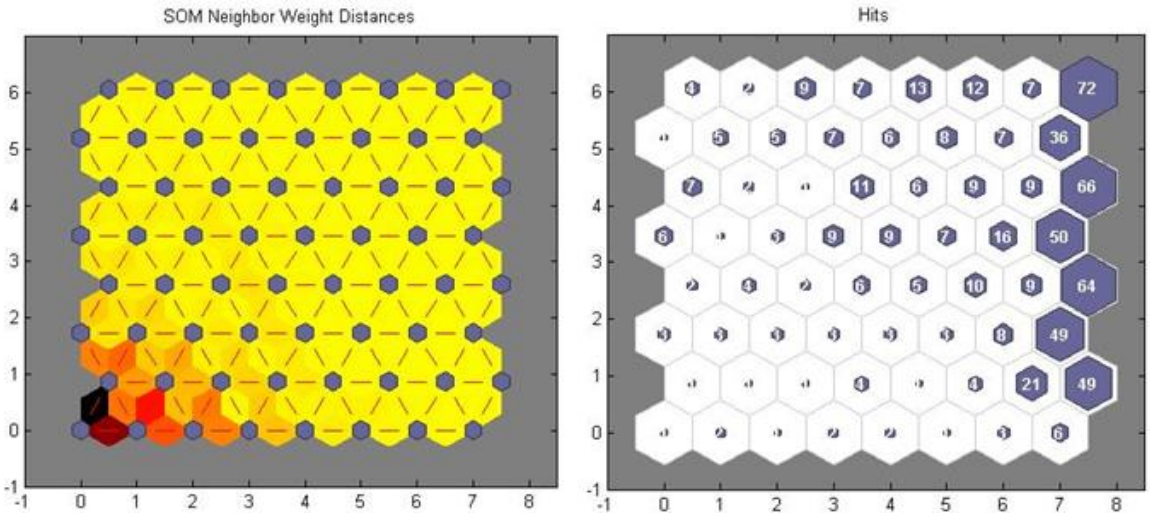


Figure 5 SOM weight distances and Hits of SOM – credit approval dataset

It is harder to identify the number of clusters in this case. We chose only 2 clusters according to graphical interpretation of Kohonen SOM network. It is highly probable that one of the clusters represents the group known as "good payers" and the second cluster represents "debtors". We could deal with the problem of single scorecard for each of mentioned group but it is beyond the scope of this paper. It is impossible to plot data points with respect to their 15 attributes (without reduction of dimensionality). We choose another way how to verify the quality of obtained solution - pair wise comparison in 2 random attributes. The comparison of 3 random situations is captured in Figure 6.

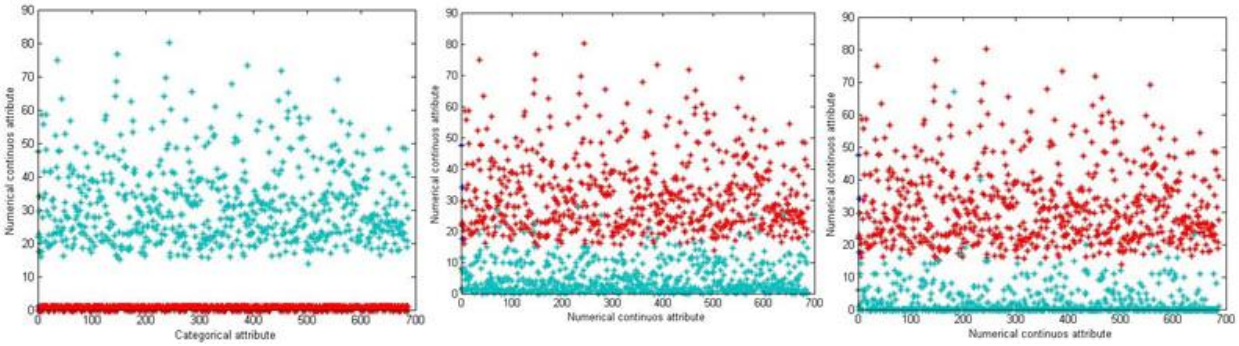


Figure 6: Attribute 1 and 2, 2 and 3, 2 and 10

Comparison for each of 2 attributes (we do not provide all of them in the interest of clarity) provides very satisfactory results and proposed algorithm seems like an effective and fast tool for clustering of big data. The biggest slowdown occurs in the finding of centroid of each cluster but proposed algorithm is faster than SOM and K-means algorithm used in isolation.

4 Conclusion

Quality of the obtained solution and primarily speed to obtain solution are definitely higher than the results obtained without combining these two approaches. There is clearly visible another option how to improve and accelerate the calculation as part of the modification of our proposed and already modified algorithm. According to the proposed algorithm it would be appropriate to implement the algorithm which is able to obtain initial data positions of the centroid of each cluster. So the final algorithm would converge to the optimal solution quickly and the whole calculation should be accelerated. The main candidates for mentioned algorithm, which would face the described problem, are genetic algorithms (or modifications of them) and tabu search. The algorithm proposed in this work will be the basis for further possible improvements and is a good springboard for further research.

Acknowledgements

Supported by the grant No. 58/2014, IG403024 of the Internal Grant Agency.

References

[1] Everitt, B. S., Landau S., Leese M. and Stahl D.: *Cluster Analysis*. 5th Edition. John Wiley & Sons, London 2011.
[2] Han, J. and Kamber, M.: *Data Mining: Concepts and Techniques.* Morgan Kaufmann Publishers, San Francisco, 2001.
[3] MacQueen, J. B.: Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1:281-297, 1967.
[4] Mays, E.*: Credit risk modeling: design and application*. GPCo, Chicago, Illinois, 1998.
[5] Mays, E.: *Credit Scoring for Risk Managers. The Handbook for Lenders*. Thomson South-Western, Mason, Ohio, 2004.
[6] Mishra M.: Kohonen Self Organizing Map with Modified K-meansclustering For High Dimensional Data Set. *International Journal of Applied Information Systems (IJAIS)*, Foundation of Computer Science FCS, New York, USA, Volume 2 – No. 3, May 2012, 34-39.
[7] SAS Institute Inc.: *SAS Enterprise Miner help and documentation, SOM/Kohonen node*. SAS Institute, Cary, United States of America, 2007.
[8] Thomas L. C.: *Credit Scoring and its Applications*. Siam, Philadelphia, 2002.
[9] Ţiţan, E. and Tudor, A.: Conceptual and statistical issues regarding the probability of default and modelling default risk. *Database Systems Journal*, Vol. II, No. 1/2011, 13-22.
[10] Tudor A. I. and Bara A., Andrei (Dragomir) E.: Clustering Analysis for Credit Default Probabilities in a Retail Bank Portfolio. *Database Systems Journal*, Vol. III, No. 2/2012, 23-30.
[11] Tudor, A., Bara, A. and Botha, I.: Solutions for analyzing CRM systems - data mining algorithms. *International Journal of Computers*, Issue 4, Volume 5.
[12] http://en.wikipedia.org/wiki/Cluster_analysis, version from 5/2/2014
[13] http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html, version from 5/2/2014

Mathematical and economic model of municipal waste management using Maple

Jana Soukopová[1], Jiří Kalina[2], Jiří Hřebíček[3]

Abstract. The paper discusses the mathematical and economical model of municipal waste management implemented in Maple system. It has been developed for the real needs of the decision support of the Ministry of Environment of the Czech Republic and regional governments. This mathematical and economical model is designed as a universal tool for prognosis of generation and treatment of municipal waste management and regional decision making. The paper introduces the proposal of an economic model of efficiency of municipal expenses. There is discussed the efficiency of regional municipal waste management system in the sample of 1 300 municipalities of the Czech Republic, reflecting the general structure of municipalities in the Czech Republic by the number of inhabitants. It depends on citizen awareness to sort recyclable household waste, an appropriate capacity of waste collection yards, an absorptive capacity of waste treatment plants and dependency mainly on the waste market prices. Important role plays financing purchase of the energy recovery from renewable sources, a use the material output of waste treatment plants and the real economic conditions of the Czech Republic.

Keywords: municipal waste, management, model, system, Maple.

JEL Classification: C290
AMS Classification: 65C20

1 Introduction

Municipal waste management (MWM) modelling is a complex process that requires a lot of data and information from various sources such as socioeconomic factors on waste generation forecasts, legislation and national Plan of waste management framework for treatment forecasts etc. We analyzed several approaches of the development of mathematical and economic models in this area [1-12] and introduce following simplification.

Let us consider that all components of municipal wastes (HW - household waste and NHW – non-household waste, i.e. similar commercial, industrial and institutional wastes) including separately collected fractions are defined unique in the European Union (EU). They are defined by the Commission Decision 2000/532/EC (European Waste Catalogue) and Annex III to Directive 2008/98/EC on waste (Waste Framework Directive).

Let the structure of our developed models issues from five aggregated municipal waste (MW) streams: *municipal waste* (MW); *mixed municipal waste* (MMW); *biodegradable municipal waste* (BMW); *material recyclable or recoverable components of municipal waste* (RMW); *hazardous components of municipal waste* (HMW). Weights for aggregation of BMW and RMW streams follow the waste codes in brackets. Weights for aggregation of MW, HMW are equal one. The special role in our models plays the composition of MMW, which consists from components: *paper* (21,1%), *plastics* (13,4%), *glass* (4,1%), *metal* (2,6%), BMW (48%) and *rest* , where percentages in brackets are results of long time research in cities of the Czech Republic [2], [6].

Moreover, there are several waste components from MW, which could be classified in more than one of the waste streams BMW, RMW and MMW, see Figure 1. For example, plastics fall into both RMW and in MMW; waste paper is included in all three groups BMW, RMW and MMW etc.

We analyzed modelling municipal solid waste generation [1], [3-4], [9-10] and European Reference Model on Municipal Waste Management [14] and modified these modeling MW generation with our approach and developed own more complex model of MW generation and treatment [7-8], [11-12]. This complex model consists of two sub-models: *sub-model of MW generation*; *sub-model of MW treatment*, which are introduced herein.

[1] Masaryk University, Faculty of Economics and Administration, Department of Public Economics, Lipová 41a, 602 00 Brno, soukopova@econ.muni.cz.
[2] Masaryk University, Faculty of Science, Research Centre for Toxic Compounds in the Environment, Kamenice 126/3, 625 00 Brno, kalina@recetox.muni.cz.
[3] Masaryk University, Institute of Biostatiscs and Analyses, Kamenice 126/3, 625 00 Brno, hrebicek@iba.muni.cz.

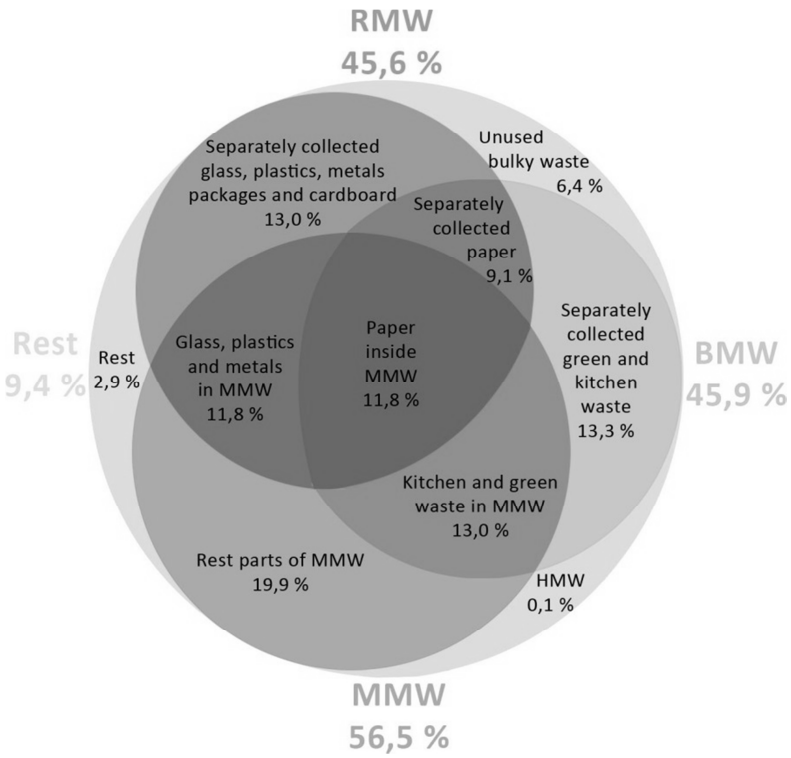


Figure 1 Composition of MW in the Czech Republic in 2012, [8]

The paper discusses the application of quantitative tools (multi-linear regression methods and material balance equilibrium method) for complex modeling of MW management (production and treatment) in the Czech Republic and forecasting MW generation and treatment within the period 2015–2024, [7-8]. These tools allow the assessment of the sustainability of environmental decisions on a national level with the focus on waste management data requirements, national strategies for waste data acquisition, management and processing in a similar way as was done in [9].

2 Sub-model of municipal waste generation

We developed and implemented the sub-model of municipal waste generation [7-8], [11] in the last two years, which is based on the principle of multidimensional linear regression like as in [1], [3-4], [9-10], [14]. We used open linked data of the individual municipalities in the Czech Republic available in e-Government systems (RISY[4], UFIS[5], ISOH[6], Ministry of the Interior[7]) and the Czech Statistical Office[8]. It did not direct the time series modelling of waste generation using extrapolation depending on time, but on the contrary to the static model of MW generation depending on time series of individual municipality parameters. In the given year, so knowledge of the parameters of the municipalities gives the prediction of MW generation and it is modeled according to the change of these parameters over time (due to the relatively long forecasting period it was necessary to work with a specific uncertainty).

The multidimensional linear sub-model of MW generation [7-8], [11] depending on a scale of municipal parameters (taking into account 20 parameters, such as population, different acreages and land use inside the municipalities, civic amenities, living standards etc.) and expected development of the parameters in a form of time series for the period 2015–2024.

Of this number of municipal parameters, 20 predictors (i.e. independent variables of the model) of the multidimensional model in total were created for each of the 6 245 municipalities of the Czech Republic by different (but only meaningful) products of type [7], [11]:

$$P_j = \prod_{i \in I_j} (p_i + c_i) \tag{1}$$

[4] http://www.risy.cz/cs
[5] http://wwwinfo.mfcr.cz/ufis/
[6] http://isoh.cenia.cz/groupisoh/
[7] http://www.mvcr.cz/clanek/statistiky-pocty-obyvatel-v-obcich.aspx
[8] http://www.czso.cz/animgraf/projekce_1950_2101/index.htm

determined by an index set I_j for this predictor (i.e. the set contains such numbers i, that $(p_i + c_i)$ should occur in the product), where P_j denotes the j-th predictor, p_i denotes i-th municipal parameter and c_i denotes an empirical constant tuned during the calibration phase of modeling.

The general form of the MW generation equation P is then given as [7]:

$$P = \sum_{j=1}^{m} (a_j P_j) + e, \tag{2}$$

where P denotes the amount of MW generation in given municipality, P_j denotes the j-th predictor, a_j denotes the j-th coefficient (searched by standard multiple linear regression method), e indicates an error estimate of municipal waste production P in a given municipality.

Submodel (1), (2) was calibrated with the data from the annual reports of the MW generation and treatment in 2011 of all municipalities of the Czech Republic using the database of the Waste information system ISOH passed from the Ministry of the Environment (MoE).

Since the predictors values are known for the whole set of municipalities in the Czech Republic (6 245) and the MW generation is known for most of them (e.g. 5 208 municipalities reported their MW generation and 1 253 municipalities reported their BMW generation in 2011), it was possible to compute the multidimensional linear model with 20 independent (predictors) and 1 dependent (production) variable, which provided the coefficients that precisely described the dependence of MW, MMW, BMW, RMW and HMW on the input parameters.

The results of time-series forecast for waste streams MW, MMW, BMW, RMW and HMW (2015-2024) generation are summarized in the Table 2.

Stream \ year	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**
MW	5.44	5.43	5.39	5.38	5.36	5.34	5.33	5.31	5.30	5.30
MMW	3.10	3.09	3.07	3.06	3.04	3.02	3.01	2.99	2.97	2.96
BMW	2.43	2.42	2.41	2.40	2.39	2.38	2.37	2.36	2.35	2.35
RMW	2.07	2.07	2.05	2.06	2.06	2.06	2.07	2.07	2.07	2.08
HMW $\times 10^3$	10.07	10.27	10.58	10.85	11.15	11.48	11.74	12.11	12.41	12.64

Table 2 The forecast generation of municipal waste streams in the Czech Republic [Tg]

3 Sub-model of municipal waste treatment

Let us consider five principal waste streams (MW, MMW, BMW, RMW and HMW) as in the waste generation sub-model; nevertheless, these waste streams are not distinct sets and their intersections change their sizes (in term of their percentages), see Figure 1, which are necessary to be included into the construction of the MW treatment sub-model. Therefore, eight distinct waste sub-streams were defined in the sub-model to enter the computation and entirely describe all the waste streams of interest [8]:

1. Amount rmw of separated recyclable waste except paper (plastics, glass, metals etc.) in MW;
2. Amount bmw of separated BMW except for paper;
3. Amount $brmw$ of separated paper and wood from MW;
4. Amount $bmmw$ of kitchen and garden waste in MMW;
5. Amount $mrmw$ of recyclable wastes in MMW except for paper;
6. Amount $bmrmw$ of paper in MMW;
7. Amount mmw of non-usable part of MMW (ashes, particulated matter, composites etc.);
8. Amount $rest$ of the rest MW (bulky waste, some hazardous parts, electro waste, … etc.).

The whole amount mwc of MW stream is defined as the sum of all above disjoined sub-streams (see Fig. 1) and similarly are defined the amount *mmwc*, *bmwc*, and *rmwc* of MMW, BMW and RMW:

$$\begin{aligned} mwc &= bmw + mmw + rmw + bmmw + brmw + mrmw + bmrmw + rest \\ mmwc &= mmw + bmmw + mrmw + bmrmw \\ bmwc &= bmw + bmmw + brmw + bmrmw \\ rmwc &= rmw + brmw + mrmw + bmrmw \end{aligned} \tag{3}$$

The amount hmw of hazardous waste stream is included in all above streams and thus we modeled it separately [8]. This simple model provides the first set of 5 equations describing the relations between the principal streams (generation of which is known) and sub-streams.

Each of the sub-stream's amounts can be further divided to five non-negative amounts coresponding to the following treatment: *material recovery, energy recovery, composting and anaerobic digestion* (AD), *landfilling* and *combustion* denoted by last five letters in the name of the model variables: *matre, enere, compo, landf, comb,* see the equation for *mwc* treatment:

$$mwc = mwcmatre + mwcenere + mwccompo + mwclandf + mwccombu \tag{4}$$

Equations for remaining sub-streams treatment are similar and we will not repeat them. The above set of equations (3), (4) led to the introduction of 60 variables in the sub-model of MW treatment and other 6 variables for dealing with HMW. This basic part of the sub-model MW treatment includes a set of 30 equations.

Some of amounts of sub-streams in above equations should be zero (such as composting of glass, metal etc. or material recovery of the unusable rest of MW). We followed waste management expert analysis and set up 10 equations, where we prescribed zero value to 10 variables. We used also the results of the sub-model MW generation for the amount of MW, MMW, BMW, RMW and HMW and obtained further 5 equations of forecasted values of *mwc, mmwc, bmwc, rmwc* and *hwc*.

The additional set of equations describes the key demands of EU and national environmental legislation and decision makers of the MoE. For example: Directive 99/31/EC on the landfill of waste (Landfill Directive) laying down the thresholds of household BMW share diverted from the flow to landfills: at least 50% in 2013 and 75% in 2020 compared to the amount of BMW landfilled in 1995 (i.e. 1.530 Tg in the Czech Republic); Waste Framework Directive sets demands specifying the amount of RMW to be recovered in 2020: the threshold is 50% here. The equations covering demands of both directives bring an element of non-linearity to the treatment sub-model, due to their nature of ratios (both consist of the proportion of a treated part of particular waste stream). E.g., the Landfill Directive criterion in 2020 could be expressed as:

$$\frac{bmwlandf + bmmwlandf + brmwlandf + bmrmwlandf}{bmw + bmmw + brmw + bmrmw} < 0.35 \tag{5}$$

Some of these amounts in above equations should be zero (such as composting of glass, metal etc. or material recovery of the unusable rest). We followed expert analysis and set up 10 equations, where we prescribed zero value to 10 variables.

The last set of equations for "refining" the sub-model and introducing 6 further variables was established in several steps as a result of discussions with waste management experts of different branches to obtain fully meaningful solution with the high precision and applicability. The results of previously made field studies of MMW composition in several Czech cities [2], [6] were included as well as the test of yield and usability of RMW from separate collections in municipalities (which yields more non-linear equations). Moreover, there was necessary to express inertia of the system during the transition between different schemes of waste management (changes in behavior of people leading to the increase of separated waste, capacity building, etc.). The detail description of equations is out of the framework of this paper and can be found in [8].

The complete set of non-linear equations of sub-model MW treatment represents 72 equations. It was solved by a computer algebra system Maple to obtain final results of different municipal waste treatment methods, which will serve as a base for Czech government decision in the next decade.

The results of sub-model MW treatment time-series forecast for 2015-2024 are summarized in the Table 3.

Streams\Year	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**
Material recovery	1.91	1.94	1.96	1.99	2.03	2.07	2.12	2.17	2.23	2.31
Composting and AD	0.37	0.43	0.49	0.54	0.60	0.65	0.70	0.75	0.80	0.85
Energy recovery	0.68	0.72	0.72	0.72	0.80	0.95	1.15	1.15	1.37	1.47
Landfilling	2.46	2.32	2.21	2.10	1.91	1.65	1.34	1.12	0.87	0.65
Combustion	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Table 3 The forecast of municipal waste treatment in the Czech Republic [Tg]

4 Economic model of waste management

We developed and implemented the economic model of waste management [7-8], [12] in the last two years which is based on the economic and financial analysis. We used open linked data of municipal expenditures on

waste management from the Ministry of Finance (MoF) of the Czech Republic automated budget systems ARIS[9] and UFIS[10] and data about population from the Czech Statistical Office.

Municipal waste management expenses (MWME) are, according to the sector budget classification, reported under items: 2122 - Collection and processing of secondary raw materials, 3721 - Collection and transport of hazardous waste, 3722 - Collection and transport of municipal waste, 3723 - Collection and transport of other waste, 3724 - Treatment and disposal of hazardous waste, 3725 - Treatment and disposal of municipal waste, 3726 - Treatment and disposal of other waste, 3727 - Waste prevention, 3728 - Monitoring of waste treatment and 3729 - Other waste treatment (remediation of old landfills and monitoring of old ecological burdens).

Let the structure of our developed models issues from five aggregated municipal waste (MW) streams: MW, MMW, BMW, RMW and HMW as in the sub-models of MW generation and treatments; nevertheless, these waste streams are not distinct sets and their intersections change their sizes (in term of their percentages), see Figure 1, which are necessary to be included into the construction of the economic sub-model. We should separate the *expenses on collection and transport - mwec* which primarily related to the amount of waste and the distance to the waste facilities (we use sub-model of municipal waste generation) and *expenses on treatment and disposal - mwet* (we use sub- model of municipal waste treatment):

$$mwe = mwec + mwet \tag{6}$$

The hazardous waste stream we modeled separately [8] and also other expenses such as waste prevention, monitoring and treatment and other waste treatment including remediation of old landfills and monitoring of old ecological burdens.

The whole amount *mwc* of MW stream is defined as the sum of all above disjoined sub-streams (see Fig. 1) and similarly are defined the amount *mmwc, bmwc,* and *rmwc* of MMW, BMW and RMW, expenses on collection and transport, treatment and disposal is function of this amount:

$$\begin{aligned} mwce &= bmwe + mmwe + rmwe + bmmwe + brmwe + mrmwe + bmrmwe + reste \\ mmwce &= mmwe + bmmwe + mrmwe + bmrmwe \\ bmwce &= bmwe + bmmwe + brmwe + bmrmwe \\ rmwce &= rmwe + brmwe + mrmwe + bmrmwe \end{aligned} \tag{7}$$

where denote:

1. *bmwe* expenses on collection, transport, treatment and disposal of *bmw* (items: 3723, 3726);
2. *mmwe* expenses on collection, transport, treatment and disposal of *mmw* (items: 3722, 3725);
3. *rmwe* expenses on collection, transport and processing of *rmw* (item 2122);
4. *bmmwe* expenses on collection, transport, treatment and disposal of *bmmw* (items: 3722, 3725);
5. *brmwe* expenses on collection, transport, treatment and disposal of *brmw* (items: 3722, 3725);
6. *mrmwe* expenses on collection, transport, treatment and disposal of *mrmw* (items: 3722, 3725);
7. *bmrmwe* expenses on collection, transport, treatment and disposal of *bmrmw* (items: 3722, 3725);
8. *reste* expenses on collection, transport, treatment and disposal of *rest* (items: 3723, 3726);

Each of the sub-stream's amounts can be further divided to five non-negative amounts corresponding to the following treatment: *material recovery, energy recovery, composting and anaerobic digestion* (AD), *landfilling* and *combustion*. On the basis of sub-models of MW generation and treatments we can model potential savings of *mwce*. We implemented the above proposed economic model in Maple and analyzed MWME with using linked open data of the MoF in the South Moravia region and national level [13].

The total waste management expenses of municipalities in the Czech Republic account for more than 4% of their total current expenditures. In the case of small municipalities with up to 500 inhabitants, it is nearly 6 % of their total expenses. The waste management expenses belong among the public expenses being the only among the environmental protection expenses (EPE) that have been growing. Currently MW treatment of municipalities are predominantly landfilling, in the South Moravian region also energy recovery. In connection with the sub-model of municipal waste treatment will therefore be the part of further research modeling potential savings of municipalities in the connection with the use of higher MW treatment hierarchy as the material recovery and composting and AD (see Table 3), using a ratio of the current MW treatment, see Table 4.

[9] http://wwwinfo.mfcr.cz/aris/
[10] http://wwwinfo.mfcr.cz/ufis/

Item of sector budget classification	2122	3721	3722	3723	3724	3725	3726	3727	3728	3729
% (total MWME)	0.17%	1.79%	75.79%	4.54%	0.29%	10.66%	0.25%	4.10%	0.08%	2.33%

Table 4 Items of budget structure and their share on total MWME 2008 – 2012

Conclusion

In the paper is presented the mathematical and economical model of municipal waste management implemented in Maple system as the result of five years research at Masaryk University in Brno. It has been used by decision makers of the Ministry of Environment of the Czech Republic for the development of the new Plan of waste management of the Czech Republic (2015-2024). This mathematical and economical model is designed as a universal tool for forecasting generation and treatment of municipal waste management and regional decision making. It can be used also in the Member States of the European Union. The paper also introduced the items of the national budget structure and their share on total municipal waste management of expenditure in 2008 – 2012.

Acknowledgements

The paper is supported by the Ministry of the Environment of the Czech Republic. Name of the project: "Development of the forecast of production and treatment of municipal waste for the Czech Republic in the period 2015–2024"..

References

[1] Beigl, P., Lebersorger, S., and Salhofer, S.: Modelling municipal solid waste generation: A review Original Research Article. *Waste Management*, **28** (2008), 200-214.
[2] Benešová, L., Černík, B., Doležalová, M., Havránková, V., Kotoulová, Z., Marešová, K., and Slavík, J.: Municipal and similar wastes (in Czech). ENZO, Praha, 2011.
[3] Emerya, A., Davies A., Griffiths, A., and Williams, K.: Environmental and economic modelling: A case study of municipal solid waste management scenarios in Wales. *Resources, Conservation and Recycling* **49** (2007), 244–263.
[4] Cherian, J., and Jacob, J.: Management Models of Municipal Solid Waste: A Review Focusing on Socio Economic Factors. *International Journal of Economics and Finance*. **4** (2012), 131-139.
[5] Hřebíček, J., and Soukopová, J.: Model of cost and price relationoships for municipal waste management of the Czech Republic, Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 7 (2011), 371-378.
[6] Hřebíček, J., Kalina, J., and Pospíchal, Z.: The development of mixed municipal waste in Brno (in Czech). *Odpadové fórum*, **6** (2012), 16-18.
[7] Hřebíček, J., Kalina, J., and Soukopová, J.: Integrated Economic Model of Waste Management: Case Study for South Moravia Region. *Acta univ. agric. et silvic. Mendel. Brun.*, **61** (2013), 917-922.
[8] Hřebíček, J., Kalina, J., and Soukopová, J.: *Developing forecasts of waste production and waste treatment in the Czech Republic in the period 2014-2024.* Part1., Part 2. 2014. Brno: Masaryk University.
[9] Hung, M.L., Ma, H., and Yang, W.: A novel sustainable decision making model for municipal solid waste management. *Waste Management*, **27** (2007), 209-219.
[10] Mazzanti, M., Montini, A., and Zoboli, R.: Municipal waste generation and socioeconomic drivers. Journal of Environment and Development, **17**(2008), 51-69.
[11] Soukopová, J. and Hřebíček, J.: Model of cost and price relationships for municipal waste management of the Czech Republic. *Acta univ. agric. et silvic. Mendel. Brun.*, **59** (2011), 371–378.
[12] Soukopová, J. and Kalina, J.: Mathematical model of economics of municipal waste management. In: *Proceedings of the 30th International Conference Mathematical Methods in Economics*. Karviná: Silesian University in Opava, School of Business Administration in Karviná, 2012, 823–829.
[13] Soukopová, J., Malý, I., Hřebíček, J. and Struk, M.: Decision Support of Waste Management Expenditures Efficiency Assessment. In: Hřebíček, J.; Schimak, G.; Kubásek, M.; Rizzoli, A.E.. IFIP Advances in Information and Communication Technology. Springer, Heidelberg, 2013. s. 651-660,
[14] Waste Model, 2012. European Reference Model on Municipal Waste Management. http://www.wastemodel.eu/ (last accessed 15.04.14.).

On Trading Price Microeconomic Estimation in Monte Carlo Simulation

Roman Šperka[1], Marek Spišák[2]

Abstract. This paper presents an experimental analysis of trading price estimation. The main method used in this partial research is the Monte Carlo simulation. The content of the paper is related to a virtual company with trading processes. The subjects of the trade are sellers and customers applying certain negotiation strategy to gain a best price for the traded commodity. The core of the negotiation is the microeconomic price estimation. Marshallian demand function and Cobb-Douglas utility function are used to count different price values. The simulation experiments should be used to support the idea that with the extension of common management information systems using a simulation method, some perspective decision support systems could be created. Their aim is to exploit the simulation to predict some Key Performance Indicators of the companies (income, profit, costs, turnover, cash flow etc.). We introduce the formal model, simulation methodology, and we depict some of the simulation results. Furthermore, we give a narrow view to the Monte Carlo simulation implementation in MATLAB. The conclusion sections will also provide a discussion of simulation results and suggest the ways in which the abovementioned concept might be applied.

Keywords: Monte Carlo, simulation, business process, trade, company, decision support.

JEL Classification: C15, C51, C53, C63, D11
AMS Classification: 68U20

1 Introduction

Business process management and simulation (BPMS) is often viewed as one of the technologies, which might help to rationalize the traditional execution of business processes in contemporary fluctuant business environment. The aim of BPMS is to lower the resources (financial, personal, time, etc.) needed for the day-to-day business praxis of companies [10, 4, 2, 8]. Regardless of the area, which is company active in, it is used to find some reserves and to lower the costs. One of the possibilities to use BPMS is an automated way to execute the business processes – Workflow Management Systems (WfMS) [9]. We used in our research more in-depth approach to the trading price negotiation to cover the seller-to-customer interaction and to use it further to simulate their behavior [13].

The motivation of this paper is to present a partial research dealing with microeconomic demand functions to be used to simulate trading price negotiation. The overall idea is based on the research of Barnett [1]. We implemented a Monte Carlo simulation [3] to support our ideas. Monte Carlo simulation method (or Monte Carlo experiments) are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results; typically one runs simulations many times over in order to obtain the distribution of an unknown probabilistic entity. The mathematical problem we dealt with was the use of some predefined functions from economic theory. We built our experimental research on a demand functions. In microeconomics, a consumer's Marshallian demand function (named after Alfred Marshall) specifies what the consumer would buy in each price and wealth situation [5], assuming it perfectly solves the utility maximization problem. Marshallian demand is sometimes called Walrasian demand (named after Léon Walras) or uncompensated demand function instead, because the original Marshallian analysis ignored wealth effects [6, 7]. We also used a Cobb-Douglas utility function and preferences saying that the quantity demanded for each commodity does not depend on income, in fact quantity demanded for each commodity is proportional to the income [15]. We have recently dealt with agent-based modeling and simulation. We used it as a core technology proving that some decision functions could be also used as a main element in business process simulation domain [12, 14, 11].

[1] Silesian University in Opava, School of Business Administration, Department of Informatics, Univerzitní nám. 1934/3, Karviná, Czech Republic, sperka@opf.slu.cz.
[2] Silesian University in Opava, School of Business Administration, Department of Informatics, Univerzitní nám. 1934/3, Karviná, Czech Republic, spisak@opf.slu.cz.

The structure of this paper is as follows. The second section introduces and explains the formal mathematical structures. The third section presents simulation results we achieved from our simulation experiments. The conclusion section sums up the results and proposes some prospective future ways to be implied in our research.

2 Formal model

For the Monte Carlo simulation the Matlab software from the Math-Works, Inc. was used. Matlab is very sufficient mathematical tool for numerical computation, visualization and for programing. In this case its ability to work with matrix and especially vectors was used. Main process of a generation for each customer is based on few vectors containing 52 values – representing 52 weeks of the year.

The customers evaluate the quotes according to the demand function by calculating its maximal price. The Marshallian demand function was derived from Cobb-Douglas utility function and represents the quantity of the traded commodity as the relationship between customer's income and the price of the demanded commodity. If the price quoted is lower than the customer's price obtained as a result of the demand function, the quote is accepted. In the opposite case, the customer rejects the quote and buys nothing.

Marshallian function specifies what would consumer buy at each specific price and income, assuming it perfectly solves utility maximization problem. For example: If there are two commodities and the specific consumer's utility function[3] is:

$$U(x_1, x_2) = x_1^{0.5} x_2^{0.5} U(x_1, x_2) = x_1^{0.5} x_2^{0.5} \tag{1}$$

Then the Marshallian demand function is a function of income and prices of commodities:

$$x(p_1, p_2, I) = \left(\frac{1}{2p_1}, \frac{1}{2p_2} \right) \tag{2}$$

Where I represents income and p_1 and p_2 are the prices of the commodities. In general, Cobb-Douglas utility function can be defined as:

$$U(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha} \tag{3}$$

The corresponding Marshallian demand function is:

$$x(p_1, p_2, I) = \left(\frac{\alpha I}{p_1}, \frac{(1-\alpha) I}{p_2} \right) \tag{4}$$

Only one commodity is calculated in the model (which is traded by the simulated company). In this case – using the Marshallian demand function there are two commodity baskets. One basket is represented by the company traded commodity and the rest represents all alternative commodities that customer can buy. So only x_1 is used supposing that the utility ratio α is known and that for the rest of commodities the utility ratio is $(1\text{-}\alpha)$. Therefore the demand function looks like this:

$$x = S \frac{\alpha I}{p} \tag{5}$$

Where X represents the amount of a commodity, α is the utility ratio, I is an income and p is the price of the commodity. Customer's decision is described by retrieving the price from the demand function. We also include the ability of the seller for the price increasing/decreasing according to his skills:

$$p = S \frac{\alpha I}{x} \tag{6}$$

This is the core formula, by which the customer decides if the quote is acceptable.

The aforementioned parameters represent a global simulation parameters set for each simulation experiment. Other global simulation parameters are:

- I – customer's income – it's normal distributed value generated at the beginning and not being changed during the generation;

[3] Ratio of the same utility, consuming the commodities – Cobb-Douglas utility function.

- α – utility ratio – a normal distributed value, which is generated for each customer each turn (week, while customers' preferences can change rapidly);
- p – commodity price;
- S – seller skills (ability to change price);
- X – an amount of commodity – normal distributed value generated, when customer decides to buy something.

The process of a generation begins for the customer. For each week he is going to decide: to buy something or not; this is uniformly distributed vector of 52 values (between *0* and *1*). If the value is *>= 0.5*, the customer decides to buy something according to the price decision (based on decision function); otherwise doesn't buy anything. After that the vectors of quantities (customer needs to decide the quantity), prices (quotes), customer incomes, customer utility ratios, and seller abilities are generated. All these numbers are derived from normal distribution with given mean and standard deviation[4].

According to the abovementioned function the quotes are evaluated and those passed are then realized.

3 Simulation results

At the start of the simulation experiments phase some parameters were set. To be more precise – 5 generations with the same parameters were done and the results were aggregated. In table 1 the complex parameterization is depicted.

PARAMETER NAME	PARAMETER VALUE
Number of Customers	10 000
Mean Quantity	50 m
Quantity S. Deviation	29
Mean Income	600 EUR
Income St. Deviation	10
Mean Utility Ratio	0.15
Utility St. Deviation	0.2
Mean Ability	1
Ability St. Deviation	0.03
Mean Sell Price	3.15 EUR
Sell Price St. Dev.	3.0
Purchase Price	0.17 EUR
Iterations count	52 weeks

Table 1 Monte Carlo simulation parameterization

The simulation was done for each customer in 52 weeks. Then the Key Performance Indicators (KPIs) were calculated. The results of the simulation are shown in 2-sided graph (Fig. 1). The results are depicted in four categories frequently used to describe the company's trading balance in general. The categories are: sold amount, income, costs, and gross profit.

[4] For particular parameterization see parameters table.

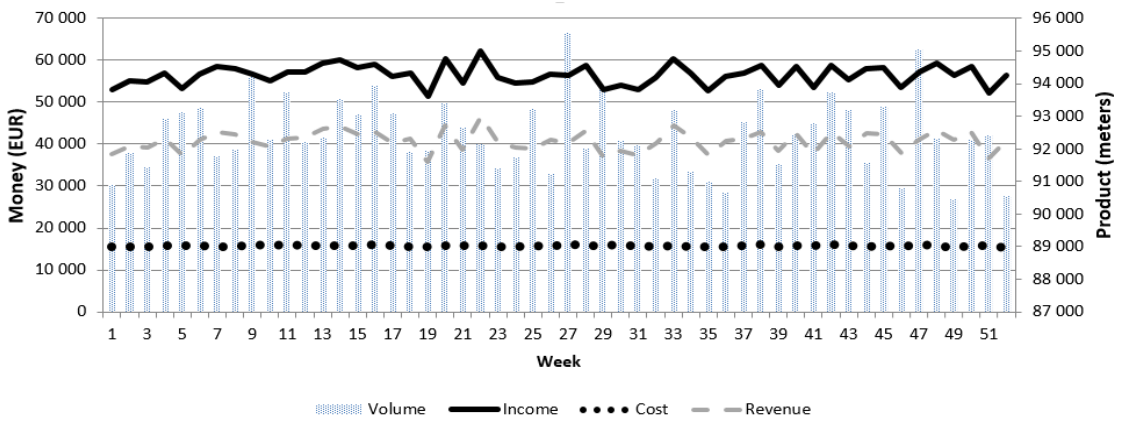


Figure 1 KPIs results for 1 year generation

The traded commodity was a UTP cable. Obviously, companies are trading with a more complex portfolio of products. We concentrated only on one product in our simplification and this was a UTP cable. Further, total gross profit was chosen as a representative KPI. Figure 2 contains the month average (AVG) of total gross profit. And figure 3 depicts a selling price for real and generated data.

Real data was provided by a Slovak anonymous company trading with computer components and supplies. The time series was discovered for the 2012 and the parameters of the simulations were set to mirror the situation on the local market of this company in that time.

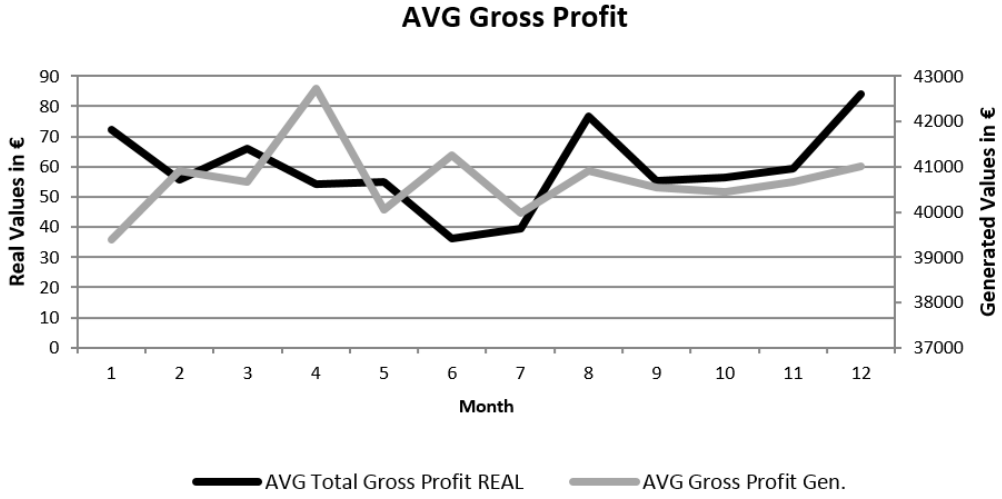


Figure 2 The generation values graph (AVG Gross Profit) – monthly

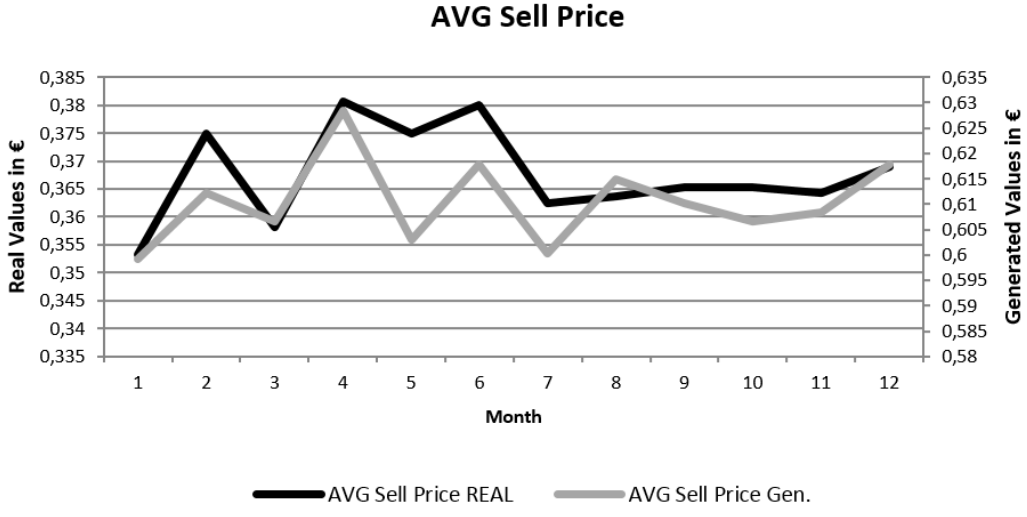


Figure 3 The generation values graph (AVG Sell Price) – monthly

To discover the correlation between the real and generated AVG month price the correlation analysis was performed. Correlation coefficient for total gross profit amount was 0.693, which represents some correlation between real and generated data. These results show that the demand functions could be used in further experiments to support the predictive purposes of prospective decision making tools based on it.

4 Conclusion

We presented some of the simulation results, which came out from the Monte Carlo simulation of trading price negotiation in a virtual trading company. We based this approach on microeconomic demand functions (Marshallian demand function and Cobb-Douglas utility function). The paper presents the formal model and simulation methodology in more in-depth way.

The results obtained show that there is some correlation between generated and real data concerning on Key Performance Indicators of a company. This shows us some promising way to use the demand function in business process simulations with the aim to implement more complex tool. This tool might be used to improve existing decision support systems of companies. This accomplishes the idea of a simulation usability in trading market environment.

We will focus on different simulation methods to support this idea in our future research. We will also compare this approach to the agent-based approach supposing that bringing some local intelligence into the seller-to-customer negotiation will proceed to even better results.

Acknowledgements

Supported by the grant “Innovation of Study Programs at Silesian University in Opava, School of Business Administration in Karvina, Reg. No. CZ.1.07/2.2.00/28.0017”.

References

[1] Barnett, M.: *Modeling & simulation in business process management*. Gensym Corporation, 2003, 6–7.
[2] Botlik, J., Botlikova, M., and Andrysek, L.: Innovative methods of regional availability analysis as a factor of competitiveness. In: *Proceedings of the 11th International Conference on Liberec Economic Forum*. Location: Sychrov, Czech Republic, 2013, 46–56.
[3] Franke, R., and Westerhoff, F.: Estimation of a structural stochastic volatility model of asset pricing. *Comput Econ* **38** (2011), 53–83.
[4] Gajdova, K., and Cieslarova, G.: Changes in selected characteristics of the EU labour market in times of economic crisis. Conference: In: *Proceedings of the19th International-Business-Information-Management-Association Conference*. Barcelona, Spain, 2011. Innovation Vision 2020: Sustainable Growth, Entrepreneurship, and Economic Development, 2012, 628–635.
[5] Marshall, A.: *Principle of economics*. 8th Ed. MacMillan, London, 1920.
[6] Mas-Colell, A., Whinston, M., and Green, J.: *Microeconomic theory*. Oxford: Oxford University Press, 1995.
[7] Pollak, R.: Conditional demand functions and consumption theory. *Quarterly Journal of Economics* **83** (1969), 60–78.
[8] Stoklasova, R.: Model of the unemployment rate in the Czech Republic. In: *Proceedings Of 30th International Conference Mathematical Methods In Economics*, PTS I AND II, 2012, 836-841.
[9] Suchánek, P.: Business Intelligence - The Standard Tool of a Modern Company. In: *Proceedings of The 6th International Scientific Symposium on Business Administration: Global Economic Crisis and Changes: Restructuring Business System: Strategic Perspectives for Local, National and Global Actors*. Karviná, Czech Republic, 2011, 123 – 132.
[10] Szarowská, I.: Effects of taxation by economic functions on economic growth in the European Union. In: *Proceedings of the 6th International Scientific Conference: Finance and the performance of Firms in Science, Education and Practice* (Jircikova, E., Knapkova, A., Pastuszkova, E., eds.) Zlin: Tomas Bata University, Czech Republic, 2013, 746–758.
[11] Šperka, R.: Application of a simulation framework for decision support systems. *Mitteilungen Klosterneuburg* **64** (2014), Hoehere Bundeslehranstalt und Bundesamt fuer Wein- und Obstbau, Klosterneuburg, Austria, Issue 1.
[12] Šperka, R., and Spišák, M.: Transaction costs influence on the stability of financial market: agent-based simulation. *Journal of Business Economics and Management* **14** (2013), S1-S12. Taylor & Francis, London, United Kingdom, Supplement 1.
[13] Šperka, R., and Spišák, M.: Microeconomic demand functions implementation in java experiments. In: *Proc. Advances in Intelligent Systems and Computing*, Volume 296, Agent and Multi-Agent Systems.

Technologies and Applications. 8th KES International Conference, KES AMSTA 2014, Chania, Greece. Berlin Heidelberg: Springer International Publishing, Germany, 2014, 183–192.
[14] Šperka, R., Spišák, M., Slaninová, K., Martinovič, J., and Dráždilová, P.: Control loop model of virtual company in bpm simulation. In: *Proceedings Advances in Intelligent Systems and Computing. Soft Computing Models in Industrial and Environmental Applications*. Berlin Heidelberg: Springer-Verlag, Germany, Volume 188, 2013, 515–524.
[15] Varian, H. R.: *Microeconomic Analysis*. Third Edition. W.W. Norton & Company, New York, Chapters 7, 8 and 9, 1992.

The analysis of the limit of obligatory offer for chosen sectors in Poland with the use of cooperative games

Anna Sroczyńska - Baron[1]

Abstract. Nowadays, taking over companies are becoming increasingly common in Europe. European Union, globalization and the competition increase make the taking over process necessary for existence and development of companies. Of course, one cannot buy every number of shares without any limitations. There are legal regulations relating to takeover bids and securities sale in each Member State. They should be according to European Parliament and Council Directive 2004/25/EC of 21 April on takeover bids. The Directive establishes a framework of general requirements which Member States are to implement through more detailed rules in accordance with their national systems and cultural contexts. In Poland, the public offering act from 2005 met with many negative opinions. Ministry of Finance has been working on the amendment of the Act on a public offer to accommodate it to the low generally accepted in Europe. There is proposed two thresholds: 33% and 66% of securities for mandatory bids for all other holdings in both cases. However, one should consider that there are usually widely dispersed shareholders in countries with the threshold at the level of one – third. It seems to be not so obvious in Poland.

In this work the structure of shareholders of chosen sectors in Poland is examined. The tools of cooperative games are used for consideration. The project of amendment of public offering act from 2005 in Poland is discussed. There is shown the possible solution of the problem of obligatory bids based on the theory of cooperative games.

Keywords: Banzhaf index, cooperative games, takeover.

JEL Classification: C71, G1, O16
AMS Classification: 91A80

1 Introduction

The history of taking over is mostly connected with American market. European markets used to be characterized by many factors which did not favor overtaking processes. The role of workers and clients was prominent, there was also a fear of taking over companies by foreign capital because of national business and strong general criticism on the grounds of ethic aspects [9]. However, changes proceeding in this part of the world –European Union forming, globalization and the growth in competitiveness– made the takeover processes necessary to keep for existence and development of companies. Bigger activity connected with hostile takeovers has recently been noticed on Polish market as well. First hostile takeover was observed in Poland in 2008 – Kruk by Wólczanka&Vistula company. Since this moment, there have been observed a lot of attempts on stock exchange in Warsaw for example Emparia by Eurocash in 2011, coalmine Bogdanka by concern NWR in 2010 or Signity by Asseco and Polfa Kutno by Concern Recordati. Of course, one cannot buy every number of shares without any limitations. Each Member State has got legalities which should conform with directive of the European Parliament 2004/25/WE. The Directive establishes a framework of general requirements which Member States are to implement through more detailed rules in accordance with national systems and cultural contexts. In Poland, the public offering act from 2005 is in force. European companies have got better legal protections toward takeover bids and concurrently there are protected the interests of small shareholders. The directive started a great discussion in Europe [3]. In Poland, the public offering act from 2005 is in force. The rules are criticized inter alia in view of a fixed threshold of mandatory bid for all other shares at the high level of 66%. The act was established according to the Directive but unfortunately did not satisfy the assumptions of defensive action for minority. Ministry of Finance has been working on the amendment of the Act on a public offer. However, the draft amendment of the Act in Poland met a great discussion and criticism. The project is similar to the law of many other Member States but met with many negative opinions in Poland [8]. It is said that there are usually widely dispersed shareholders in countries with the threshold at the proposed level what is not so obvious in Poland.

[1] University of Economics in Katowice, Department of Applied Mathematics, ul. 1 Maja 50, 40 – 287 Katowice, kozak@caprisoft.com.pl

The main aim of this work is the analysis of the shareholders structure of chosen sectors of companies traded on the Stock Exchange in Warsaw based on the theory of games and the discussion of proposed level of threshold of mandatory bid in the draft amendment.

2 The Act on a public offer from 2005 in Poland

Legal regulations connected with takeovers should respect European Parliament and Council Directive 2004/25/EC of 21 April on takeover bids. But the Directive only establishes a framework of general requirements which Member States are to implement through more detailed rules. Till now, in Poland an offeror is obliged to keep the following legal rules applicable to the conduct of bids according to the Act of 29 July 2005 on a public offer:

- the offeror is required to make a bid of the purchase of shares representing not less than 66% of all in the offeree company when the offeror is going to hold shares representing at least 33% of them in that company,
- the offeror is required to make a bid of the purchase of all other shares of the offeree company when the offeror is going to hold shares representing at least 66% of them in that company.

The rules are criticized inter alia in view of two fixed thresholds of mandatory bids: 33% and 66%. The limit 33% often confers control in a company but it allows to launch a bid only till 66% of shares. There are fears for protection of minority shareholders – not all of holders of remaining securities are able to require the offeror to buy their securities at a fair price and to withdraw from the investment in such a situation. On the other hand, the obligation to launch a bid for all other holdings after the threshold of 66% of securities applies when the takeover is in principle conducted. The majority shareholder can stop the acquiring at the level of 65% of securities which certainly gives him the control of the offeree company without the obligation of a bid for the rest of securities. It is visible that defensive action for minority shareholders is too weak and do not satisfy the Directive in Poland. Although the Directive does provide for arrangements in the area of the protection of minority shareholders, it leaves it up to Member States themselves to establish the threshold of mandatory bid. Ministry of Finance has been working on the amendment of the Act on a public offer. There are still proposed two thresholds: 33% and 66% of securities for mandatory bids but for all other holdings in both cases. An offeror would be obliged to keep the following legal rules applicable to the conduct of bids according to the amendment of the Act of 29 July 2005 on a public offer:

- the offeror would be required to make a bid of the purchase of all other shares of the offeree company when the offeror is going to hold shares representing at least 33% of them in that company,
- the offeror would be required to make a bid of the purchase of all other shares of the offeree company when the offeror is going to hold shares representing at least 66% of them in that company.

The proposition seems to be rational and is commonly used in Europe. However, in Poland there is a fear that the threshold at the level of one – third shares is not proper one. It may cause pull outing the companies from the stock exchange even against first intentions of offerors and may influence the market's liquidity. Markets with the threshold at the level of 33% of shares are usually characterized by great participation of free float, what is controversial in Poland.

3 Methodology

The theory of games is a useful tool for proper decisions in the world of economics especially when we are not certain of how the other subjects are going to act. It seems to be a suitable tool to use on the stock exchange [10]. One of its practical applications connected with the stock exchange is using n – person cooperative games for the analysis of hostile takeovers. The theory n – person is completely different from the theory of two - person game because of possibilities of coalition forming. It often makes it possible to increase the expected payoff for the player. Let S be a subset of players who made decision to form a coalition. It means that they will decide about individual actions so that the sum of payoff for all members of coalition would be possibly the best. It is assumed that the rest of players form the coalition S'. Then the game comes down to two – person game: coalition S contra coalition S'. In this work the Shapley's method is used to solve it. He examined the problem of possibilities of game estimates for each player. He formulated the idea of the value of the game. It is defined as the weighted sum of profit growths, which one player gives to all coalitions he belongs to. Power index is a particular value of games for simple games. Simple games have a characteristic function which takes only two values: 0 or 1 – it means, that the coalition is a wining one or not. One of the most popular indices is Banzhaf index. Banzhaf index is most often taken to examine the shareholders structure [7]. It measures power as an influence. The value which is gained by the coalition is the possibility of control. This kind of index also ignores the order of players attachment to coalitions. It is calculated according to the following formula.

$$\beta_i(v) = \frac{c_i}{\sum_{j \in I_n} c_j} \tag{1}$$

where c_i is a number of coalitions with the player i as a decisive player (it means the coalition with this player is a winning one, without him is a loosing one), v – characteristic function of the game, In - the set of n players. Banzhaf index assigns a part of whole winnings to each player. The part is proportional to number of coalitions for which the player is a decisive one.

In this work the power index was used to check the situation of a company – offeror and the threshold of 33% for obligatory offer. The model of large games with an outcome majority (half of shares, which are possessed only by main players) was used to analyze it [6]. It was assumed that there were m main players and infinitive number of small players who did not have any influence individually. The tools of the theory of games, as for example grey systems in [1],[2] could be effective alternatives for traditional methods of market analysis. There are many popular tools to evaluate the companies, for example TMAI [12] and others [13], but to evaluate shareholders structure of those companies Banzhaf index seems to be proper one [11].

4 The researches

4.1 Data

The analysis of chosen sectors of stock exchange in Warsaw was conducted during the researches. Two sectors were chosen for examination: telecommunication and computer. There are said to be the most exposed to the risk of hostile takeovers. These sectors are characterized by the biggest free – float. What is more, the companies from the telecommunication sector have worse results that the market – index WIG – telecom had the value of -14,85% for the last year [5]. There are 6 companies from telecommunication sector and 19 companies of IT sector traded on the Stock Exchange in Warsaw. There are usually four main shareholders and free float develops at the level of 40%. There is only 6 companies where the majority shareholder exists (the shareholder who posses more than 50% of shares). But most of them are really exposed to hostile takeovers because of minority control. The level of free float of companies with minority control from these two sectors is presented on Fig. 1.

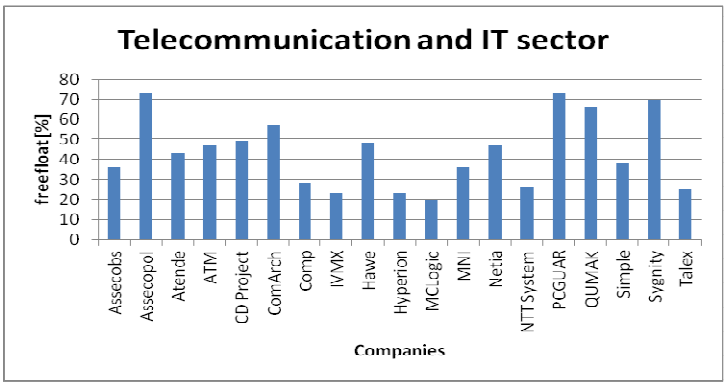


Figure 1 Free float of telecommunication and IT sector Warsaw Stock Exchange for companies with minority control 12.2013 in [%], source: own work based on http://www.stockwatch.pl

4.2 The run of researches

There is proposed the threshold of mandatory bids at the level of 33% according to the amendment of the Act on a public offer in Poland. So the analysis of the situation of examined companies was conducted during hypothetical attempt of purchase 33% of shares by the aggressor. Will the free float be protected in this situation and concurrently won't the threshold become a barrier and will cause phasing out by companies and loosing the fluid? During the researches the Banzhaf index was used to check the situation of main shareholders of companies presented in Fig 1. The model of large games with outcome majority was used during the calculations. It was assumed that the aggressor was a player B. The obtained results are presented in Table 1. There are the values of Banzhaf index for real main shareholders in situation when the player B would be the theoretical invader with 33% of shares.

	β_A	β_B	β_C	β_D	β_E	B_F	B_G	Outcome majority
Assecobs	1	0	0	-	-	-	-	43
Assecopol	0	1	0	-	-	-	-	26
Atende	0,09	0,64	0,09	0,09	0,09	0	-	41
ATM	0,17	0,5	0,17	0,17	0	-	-	39
CD Projekt	0,05	0,75	0,05	0,05	0,05	0,05	-	36
ComArch	0	1	-	-	-	-	-	33
Comop	0,12	0,5	0,08	0,08	0,08	0,08	0,08	46
IVMX	0,39	0,23	0,23	0,08	0,08	-	-	48
Hawe	0,2	0,44	0,2	0,12	0,04	-	-	39
Hyperion	0,25	0,28	0,12	0,12	0,08	0,08	0,08	49
MCLogic	0,25	0,28	0,12	0,12	0,08	0,08	0,08	47
MNI	0,33	0,33	0,33	-	-	-	-	42
Netia	0,17	0,5	0,17	0,17	-	-	-	35
NTT System	0,33	0,33	0,33	0	0	-	-	42
PCGUARD	0	1	0	0	-	-	-	27
QUMAK	0	1	0	0	0	-	-	30
SIMPLE	0,17	0,5	0,17	0,17	-	-	-	37
Sygnity	0	1	0	0	-	-	-	28
TALEX	0,33	0,33	0,33	-	-	-	-	42

Table 1 Banhaf Index for companies from telecommunication and IT sectors with minority control during hypothetical attempt of purchase 33% of shares by the player B, source: own computation

It was showed that the threshold of 33% allowed to gain the control in a company only in 5 occurrences: company Sygnity, Qumak, PCGuard, ComaArch and Assecopol. These are the companies characterized by the biggest free float – more than 55%. For all other companies the threshold seems to be too low. Investors are required to buy all other shares without any control in a company. In fact, it may really influence the market's liquidity. Next part of the researches is connected with the attempt to estimate the proper threshold of obligatory bid for these two sectors. It was examined how many shares would the potential offeror need to gain the minority control in the companies from telecommunication and IT sectors? It was checked how many shares would give the value of index 1 for the player B. The results are presented in Figure 2.

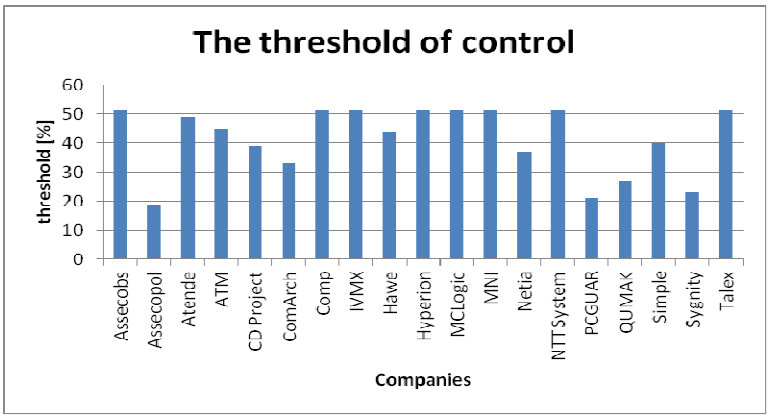


Figure 2 The level of minority control for companies of telecommunication and IT sector Warsaw Stock Exchange 12.2013 in [%], source: own work

The researchers showed that the level of minority control is much higher for the companies from examined sectors than in the draft amendment – the average percentage is at the level of 41,3%. It means that the threshold of 33% would not work properly. Offerors would have to buy all other shares without any control in the company. On the other hand the examination of 5 companies where the threshold of 33% was enough showed that the truth was that it could be much lower. The average threshold of minority control for these 5 companies was 24,6%. So there was no protection of small shareholders as well. The offeror could buy much less than 33% of shares, gained the control in the company and would have to buy no more shares. Only for two companies: ComArch and Qumak it is honest to say that the threshold at the level of one – third is correct.

The researchers showed that the threshold of 33% seems to be proper in countries, where the companies with dispersion of shareholder structure are predominate (it worked rather in good way for companies with big freefloat). However, in Poland and other countries with opposed tendency the solution should be different.

5 Conclusions

In Poland the problem of the threshold fixed by the Act of 29 July 2005 on a public offer has been discussing for a few years. Legislations were conducted according to the Directive of the European Parliament 2004/25/WE. Unfortunately, it seems that it did not meet with objectives of the Directive. The draft amendment of the Act is similar to the law of many Member States but it met also with great criticism because of the nature of Polish market. In this work the companies from telecommunication and IT sectors were examined. The researches showed that the participation of free float was at the level of 40% and the threshold of 33% for obligatory offer did not work properly for these companies. The participation of small shareholders is not enough for a level of one – third for the threshold. The offerors are often required to make an obligatory offer without the position of dominator and concurrently there are not protected the interests of small investors in many situations because the company – offeror is not obliged to make bid for other shares after acquiring the control. So it worked in a proper way only in 6% of companies. What is important, the companies examined in the work belonged to sectors characterized by the biggest free float at the Stock Exchange in Warsaw. The situation in other sectors with smaller free float would be even worse. It seems to be worth the discussion to increase the threshold of obligatory bid in Poland or to use just the power index to fix the threshold . It seems to be a better tool than percentages and universal tool analyzing the structure of the power, insensitive to shareholder structure, especially with the model of large games presented in this work. Fixing one level of this index as the threshold of obligatory bid would allow to standardize the situations of companies – offerors at the moment of excess of it and would better protect the interests of small shareholders.

The problem of the measurement of influences seems to be always actual and very important. Good solutions of this problem in one area could help to solve similar problems in other, which the European Union meets, for example the rotation voting system recently adopted by the European Central Bank [4].

References

[1] Barczak, S.: Zastosowania modeli szarych do prognozowania procesów s kształtnych na przykładzie prognozy na rok akademicki 2011/2012 liczby studentów wszystkich typów studiów w Polsce. In*: Zastosowanie metod ilościowych w zarządzaniu ryzykiem w działalności inwestycyjnej* (Tworek, P., and Barczak, A., eds.). Wyd. UE w Katowicach., Katowice, 2013, 482-496.

[2] Barczak, S.: Zastosowanie teorii szarych systemów do przewidywania przyszłych ofert składanych na aukcjach pierwszej ceny poprzez pryzmat modelu szarego GM(1,1). In*: Studia Ekonomiczne. Innowacje w finansach i ubezpieczeniach – metody matematyczne i informatyczne* (Mika, J., and Dziwok, E., eds.). Wyd. UE w Katowicach., Katowice, 2013, Vol. 146, 7-18.

[3] Clarke, B.: The takeover directive: Is a little regulation better than no regulation?. *European Law Journal* **15(2)** (2009), 174-197.

[4] Di Giannatale, P., & Passarelli, F.: Voting chances instead of voting weights. *Mathematical Social Sciences* **65(3)** (2013), 164-173.

[5] Fura M.: Prób wrogich przejęć w Polsce będzie coraz więcej, Retrieved from *ww.obserwatofinansowy.pl*, 2013.

[6] Gambarelli, G., & Owen, G.: The coming of game theory. *Theory and Decision* **56** (2004), 1-18.

[7] Leech, D.: An empirical comparison of the performance of classical power indices. *Political Studies* **50(1)** (2002), 1-22.

[8] Obrycka, A.: O nieobowiązkowym obowiązku wezwań i projekcie założeń nowelizacji ustawy o ofercie publicznej. *Biuletyn Koła Naukowego Prawa Spółek* **2** (2012), 2-8.

[9] Prokop, J.: *Przyczynek do teorii przejmowania kontroli nad spółkami akcyjnymi*. Wydawnictwo SGH, Warszawa, 2001.

[10] Sroczyńska-Baron, A.: The choice of portfolio based on the theory of cooperative games. In: *European Financial Systems 2013. Proceedings of the 10th International Scientific Conference* (O. Deev, V. Ksjurova & J. Krajicek. eds.), Masaryk University, Brno, 2013, 305-311.
[11] Sroczyńska-Baron, A.: The analysis of the process of taking over companies based on the theory of games. In: *The 7th International Days of Statistics and Economics, Conference Proceedings* (Löster, T., and Pavelka, T., eds.). Libuše Macáková, MELANDRIUM, Prague, 2013, 1324-1333.
[12] Węgrzyn, T.: Stock selection based on financial ratios on the Warsaw Stock Exchange. Analysis between 2001 and 2010. In *European Financial Systems 2013. Proceedings of the 10th International Scientific Conference*, (O. Deev, V. Ksjurova & J. Krajicek. eds.), Masaryk University, Brno, 2013, pp. 356-361.
[13] Węgrzyn, T.: Stock Selection on the Warsaw Stock Exchange Financial Ratios or Profitability Ratios. Analysis between 2001 and 2011. In: *The 7th International Days of Statistics and Economics, Conference Proceedings* (Löster, T., and Pavelka, T., eds.). Libuše Macáková, MELANDRIUM, Prague, 2013, 1554-1564.

Convergence to Monetary Equilibrium in Search Model

Rostislav Staněk[1]

Abstract. Monetary equilibria are welfare superior comparing to barter equilibria in models with decentralized trade, such as search models. However, monetary equilibrium arises only if agents share the belief that money is accepted by other agents. The paper explores the question whether agents can spontaneously coordinate on such a belief in case of fiat money. Formation of monetary equilibrium is studied in the monetary search model, which is similar to Kiyotaki-Wright model. The model is characterized by decentralized trading where agents have to decide whether to accept fiat money or not. Each agent has some belief about money acceptability and chooses optimal action with respect to this belief. The belief is updated after each trade. To allow for non-equilibrium dynamics the model is solved by means of agent based simulation. The paper shows what conditions have to be met in order for monetary equilibrium to arise. The effects of network type, total number of agents in society and initial conditions on the acceptability of fiat money are examined.

Keywords: monetary equilibrium, search model, convergence, agent-based model

JEL classification: D83, C73
AMS classification: 91A, 91D30

1 Introduction

One of the fundamental questions in monetary economics is why fiat money has value: Why do rational agents trade real resources for intrinsically worthless pieces of paper? Monetary economists have long understood that part of the explanation relates to the superiority of a monetary equilibrium over a barter one. Search models of money such as [4] and [6] shows that fiat money is essential to facilitate exchange in the model with decentralized trade. This means that the model have a pareto-superior equilibrium when money is accepted. However, it is one thing to explain that equilibrium with monetary exchange exists, and quite another to explain how monetary equilibrium emerges.

This paper provides a contribution to the literature on emergence of money in search models. It focuses on the question, whether there is any dynamic process which leads to monetary equilibrium. Existing contributions to this question studies convergence to monetary equilibrium from barter exchange. Luo [3] uses an evolutionary approach to explain emergence of money. He shows that initial freqeuncy of strategies and storage costs of commodity are crucial determinants for emergence of money. Iwai [2] shows that monetary equilibria are stable but large disturbances are needed for monetary equilibrium to emerge. Hence, he concludes that there is no natural tendency for the evolution from barter to monetary economy. Duffy [1] uses agent based simulation to explore how unequal distribution of player's strategies determines the speed of convergence to monetary equilibrium. All of these papers consider emergence of money from the most acceptable commodity, which means that each unit of money can be directly consumed by some agent. This assumption contradicts with definition of money in search models [4] [6]. The aim of this paper is to show whether monetary equilibrium can emerge in the search model with fiat money and which factors are important for emergence of monetary equilibrium. In accordance with search models, I suppose that money cannot be consumed and does not bring any direct utility to any agent. Therefore, money can be accepted by any agent only if other agents are willing to accept money as well. The model presented in this paper is solved by means of agent-based simulation. Besides monetary

[1]Masaryk University, Faculty of Economics and Administration, Lipová 41a, Brno, Czech Republic, rostanek@mail.muni.cz

search literature, the paper follows also literature investigating exchange on fixed networks (see e.g. [5], [7])

2 Model

2.1 Trading technology

The model consists of three components: trading technology, learning process and network topology. First, we describe the trading technology. Trading technology is characterized by imperfect monitoring of past actions and anonymity and absence of double coincidence of wants. In particular, consider an economy with M infinitely-lived agents and N perishable and indivisible types of goods, where $N \geq 3$. There are N types of agents. Denote M_i the number of agents of type i. It is clear that $\sum_{i=1}^{n} M_i = M$. An agent of type i produces only good i and consumes some other type of good that is randomly chosen in each period. Time is discrete and infinite, indexed by $t = 1, 2, 3...$. At each date, each agent meets one other agent that produces the good that she demands. This model environment is the same as in the baseline random matching model of [4] with the difference that there is finitely many agents in our environment, whereas there is a continuum of agents in [4].

It is clear that money is necessary to facilitate exchange in the model. Suppose therefore that some agents are endowed with one indivisible unit of money. It is obvious that the agent has a possibility of choice among some actions only if she is a producer. Then she can choose if she will deliver a good to his trading partner. For simplicity assume that each agent is able to produce only one unit of good. The action set of any agent-producer i is therefore $a_{t,i} = \{0, 1\}$, where 1 denotes producing and delivering of one unit of good to his trading partner in exchange for money and 0 denotes no trading. The general utility function of agent i for one period is $u_i(a_{i+1}) - a_i$, where a_{i+1} is the amount of good consumed and a_i is the amount of good produced. Suppose that the utility function such that $u(0) = 0$ and $u_i(1)$ is uniformly distributed on the interval $[1, \bar{u}]$. It holds that $u_i \geq 1$, i.e. utility that the agent yields from consumption of one unit is greater than production cost. Money does not bring any direct utility to the agent. So, the agent is willing to accept money only if he believes that future trading partners will accept money as well.

2.2 Learning

The agents update their belief by algorithm known as fictious play. Fictious play involves two parts. Each agent forms expectations about the current play of the opponent in a simple and empiricist manner, she associates subjective probabilities assigned to strategies of future trading partner with respective frequencies of these strategies in past play. Given the expectations formed in this fashion at each point in time, each player reacts optimally.

To define the learning process formally, denote by $\rho_{is}(t)$ the weight associated by agent i at time t to the strategy $s \in \{0, 1\}$. To start the process, players attribute initial weights to the strategies $\{0, 1\}$. Depending on the observed action of the trading partner $a_j(t) \in \{0, 1\}$ the weights are adjusted as follows

$$\rho_{is}(t+1) = \begin{cases} \rho_{is}(t) + 1 & \text{if } a_j(t) = s, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

These weights are then used to form player's expectation $e(t)$ at time t in the following manner

$$e_{is}(t) = \frac{\rho_{is}(t)}{\sum_{s'} \rho_{is'}(t)} \tag{2}$$

Finally, each player chooses a strategy that belongs to the best-response correspondence given player's belief. If the agent produces good and accepts money then she can buy consumption good in the future period when she meets producer that is willing to accept money. The expected payoff that the agent receives for accepting money is given as follows

$$U(1) = -1 + \delta\, e_{i1}\, u_i + \delta^2\, (1 - e_{i1})\, e_{i1}\, u_i + \cdots = \frac{\delta\, e_{i1}\, u_i}{1 - \delta(1 - e_{i1})} - 1$$

The agent chooses the action which gives him the highest payoff. Hence, she is willing to produce one unit of good and accept money in exchange if and only if $U(1) \geq 0$.

2.3 Matching topology

The trading process in the model is decentralized. In order to simulate the structure of decentalized meetings, the agents are placed in some network structure. In each round of the model, every agent (in random order) is matched with some agent to which she has a network connection. Hence, each agent can trade only with agents to which she has a network connection. The paper considers three types of network topology: complete network, locally connected network and small world.

- In a complete network each agent is connected to every other agent. Therefore, each agent can be matched with all other agents with equal probability. The simplicity of this network makes it a good benchmark network.
- The agents are split into distinct groups. The groups are then arranged in a circle and each group overlaps its neighboring groups by sharing one agent. Thus every group has two agents who are common to different adjacent groups. The common agents can trade with members of either group while the rest of the members of a group can only trade within the group.
- Small world network lies between complete network and locally connected network. The agents are formed into a locally connected network. Then a small number of agents are randomly selected to connect with agents in other groups. There are two restrictions on these crossover connections. First, a crossover agent cannot be one of the common traders shared between adjacent groups. Second, the new connection cannot connect two groups who already have a common agent.

Each of these network is characterized by different degree of clusteering. To measure the clustering in the network, a common measure is the clustering coefficient $C \in [0, 1]$ (see [7]). The clustering coefficient of each agent is defined as the proportion of neighbors of that agent that are connected. The clustering coefficient for the whole network is given as the average of the local clustering coefficients. Obviously, clustering coefficient in the complete network is equal to 1. Clustering coefficient in the locally connected network is given as the reciprocal for the number of groups.

3 Results of simulations

The model has been simulated for number of agents $M \in \{50, 75, 100\}$ and the three above mentioned types of network. In the locally connected network and in the small world network agents were divided into 4, 6, 8 and 10 number of groups. In the small world network agents from different groups could have been connected by 2,3,4 or 5 crossover connections. The discount factor was set to 0.9; 0.92 and 0.94, i.e. $\delta \in \{0.9, 0.92, 0.94\}$. The model was simulated for three different upperbound value of utility $\bar{u} \in \{0.8, 1, 1.2\}$. The initial weight to the strategy *accept money* was set to 1, i.e. $\rho_{i1}(0) = 1$, while the initial weight to the strategy *do not accept money* is a member of the set $\{1, \ldots, 9\}$. The model was simulated thirty times for each feasible combination of parameters and each variant. The total number of the simulation runs was 349 920. The maximal amount of simulation steps was set to 2 000. The model was simulated in NetLogo 5.0.4. Since the full parametric space of the simulation is huge, only the most salient stylized facts will be presented here.

3.1 Simulation outcomes

The aim of the simulations is to answer two questions: 1) What is the probability of emergence of monetary equilibrium? 2) What are the main determinants of emergence of monetary equilibrium? This section focuses on the former. The descriptive statistics of simulation outcomes are reported in table 1. We can see that agents expected in average that 79 % of other agents will accept money and 80 % of agents were willing to accept money. This numbers are quite high especially when the initial parameter setting was chosen such that the starting average agent's expectation regarding money acceptability was only 26 %. This suggests that agents have not any obvious diffculty to adopt monetary trading strategy.

	Mean	Std. deviation	Median	Max	Min
Acceptability	0.79	0.22	0.87	0.98	0.03
Acceptance ratio	0.8	0.24	0.88	1.00	0.00
Convergence	0.71	0.3	0.73	1.00	0.00
Initial acceptability	0.26	0.13	0.25	0.5	0.1
Initial acceptance ratio	0.52	0.24	0.57	0.87	0.01

Table 1 Descriptive statistics of simulation outcomes. *Acceptability* is the average expectation that money is expected. *Acceptance ratio* is the share of agents that are willing to accept money. *Convergence* is the share of groups with acceptance ratio higher than 0.9 or lower than 0.1

The simulations produce qualitatively different kinds of outcomes. There are three possible steady-states: monetary equlibrium, autarky equilibrium, and mixed steady-state. Monetary equilibrium is defined as the outcome when acceptance ratio excceds some given treshold. If not stated otherwise this treshold is set to 0.9. The simulations converged to monetary equilibrium in 61.2 % of runs. Autarky equilibrium is defined as the outcome when acceptance ratio is less than given treshold. The treshold is given as 0.1. The simulations converged to autarky equilibrium in 5.7 % of runs. In 33.1 % of runs the simulation did not converged to neither monetary nor autarky equilibrium. Since the above stated frequencies are somewhar artificial and depend on the simulated parameter space, it is instructive to decompose them. The decomposition can be found in table 2. Monetary equilibrium occurred in more than 70 % of runs when the initial share of agents accepting money is at least 20 % and the network was either complete or with small number of groups and many crossover connections. Note that complete network converged to monetary equilibrium in 20 % of runs although the initial ratio of agent's accepting money is below 20 %. On the other hand, autarky equilibrium arises only if initial ratio of agent's accepting money is below 20 %. Note also that high clustering coefficient increases the probability that the simulation converges to monetary or autarky equilibrium. The tabel 2 suggests that initial conditions and network type can be crucial factors determining emergence of monetary equilibrium.

The mixed steady-state occured most often when the clustering coefficient is low. The situation occurs because the network is very local, i.e. there are a lot of groups and with small number of crossover connections. Despite the fact that the simulation does not converged to any pure equilibrium, there is a clear tendency to create clusters. The clusters are mostly overlapping with trading groups. Half of trading groups converged to monetary or autarky equilibria in these runs.

IAR/CC	$[0, 0.1]$	$[0.1, 0.2]$	$[0.2, 0.3]$	$[0.3, 1]$	$[0, 0.1]$	$[0.1, 0.2]$	$[0.2, 0.3]$	$[0.3, 1]$
$[0, 0.2]$	41.2%	43.4%	39.7%	61.5%	1.6%	2.9%	4.8%	20.1%
$[0.2, 0.4]$	0.7%	0.6%	0.8%	1.4%	39.4%	45.7%	59.8%	71.9%
$[0.4, 0.6]$	0%	0%	0%	0.1%	57.3%	61%	66.9%	70.5%
$[0.6, 0.8]$	0%	0%	0%	0%	70.2%	73%	76.5%	78.4%
$[0.8, 1]$	0%	0%	0%	0%	88.1%	90.1%	92.1%	93%

Table 2 Relative frequency of emergence of monetary equilibrium for given clustering coefficient (CC) and selected initial accepting ratio (IAR). The left-hand side of the table shows the relative frequency of emergence of autarky equilibrium. The right-hand side of the table shows the relative frequency of emergence of monetary equilibrium.

3.2 Determinants of convergence

This section investigates what factors influence emergence monetary or autarky equilibrium. The overall probability that some equilibrium type occurs can be modeled with the logistic regression. The results are summarized in table 3, which shows determinants of probability of monetary equilibria (right-hand side) and autarky equilibria (left-hand side). The presence of a highly clustered social network increases the probability of both types of equilibria. Therefore, it is clear that this factor decreases the probability of mixed steady-state emergence. The rise in the consumption utility, discount factor and initial acceptability increases the probability of monetary equilibrium emergence and decreases the probability of autarky

equilibrium emergence. The effect of these variables could be explained by the fact that they increase the initial share of agents that are willing to accept money. Hence, the results show that monetary equilibria occurs when surplus generated by exchange is high, agents are patient enough and agents have a lot of connections to other agents.

	Estimate	Std. error	P-value	Estimate	Std. error	P-value
Constant	352	6.6	0.0000	-99	0.72	0.0000
Prior acceptability	-165	6.2	0.0000	7.8	0.1	0.0000
Utility	-30	0.7	0.0000	7.4	0.1	0.0000
Discount factor	-335	0.044	0.0000	98	0.7	0.0000
Clustering coeff.	12	0.7	0.0000	9	0.3	0.0000
	McFadden's	$R^2 = 0.68$		McFadden's	$R^2 = 0.59$	

Table 3 Binary logit model. Dependent variable is the autarky equilibrium in the left-hand part of the table and the monetary equilibrium in the right-part of the table.

Initial conditions and the type and size of the social network do not affect only the qualitative, but also the quantitative characteristics of the equilibrium. Table 4 shows the determinants of the average acceptability (left-hand side) and acceptance ratio (right-hand side). The dummy variables for local network and small world network have negative effect on both dependent variables. This result occurs because both networks have lower connectivity compared to complete network. Therefore, there is a higher probability that a group of agents that do not accept money arise in these types of network. As we have seen before, consumption utility, discount factor and initial acceptability have positive effect on the equilibrium acceptability and acceptance ratio.

	Estimate	Std. error	P-value	Estimate	Std. error	P-value
Constant	-5.54	0.043	0.0000	-5.87	0.048	0.0000
Prior acceptability	0.68	0.005	0.0000	0.68	0.006	0.0000
Utility	0.38	0.008	0.0000	0.39	0.008	0.0000
Discount factor	6.33	0.044	0.0000	6.69	0.049	0.0000
Small world	-0.04	0.002	0.0000	-0.02	0.002	0.0000
Local network	-0.04	0.002	0.0000	-0.04	0.002	0.0000
	Adjusted $R^2 = 0.62$			Adjusted $R^2 = 0.61$		

Table 4 Linear regression model. Dependent variable is average acceptability in the left-hand part of the table and share of agents accepting money in the right-part of the table. Small world and Local network are network dummy variables. The robust standard errors of the parameter estimates are reported.

4 Conclusions

Current literature show that commodity money can spontaneously arise in search models. This paper presents a simulation model of fiat money emergence. In addition to the previous literature, the simulation results show that agents can coordinate in monetary equilibrium even if money are intrinsically useless and cannot be consumed directly. Moreover, some conclusions of the previous literature [2] have to be modified. The monetary equilibrium emerges quite easily and it does not require any special initial conditions or large disturbances. Only a relatively small initial fraction of agents accepting money is sufficient for monetary equilibrium to emerge. Although fictious play constitutes relatively simple learning algorithm it ensures emergence of monetary equilibrium. This implies that even not very sophisticated agents can coordinate themselves in monetary equilibrium.

In general, the type of the resulting equilibrium, and the equilibrium expectations regarding money acceptibility are quite sensitive to the initial conditions and type of the used network. Small world network and locally connected network decreases the probability of money emergence compared to the complete network. This relationship can be explained by lower clustering coefficent of these networks. Lower connectivity prevents the spread of monetary equlibrium. This results suggest that a careful specifiacation of social network can be crucial for modeling real-world currency competition.

Acknowledgements

The paper was created as a part of the specific research project number MUNI/A/0781/2013 at Masaryk University.

References

[1] Duffy, J.: Learning to speculate: Experiments with artificial and real agents. *Journal of Economic Dynamics and Control*, **25** (2001), 295–319.

[2] Iwai, K.: The bootstrap theory of money: A search-theoretic foundation of monetary economics. *Structural Change and Economic Dynamics*, **7** (1996), 451–477

[3] Luo, G.: The evolution of money as a medium of exchange. *Journal of Economic Dynamics and Control*, **23** (1999), 415–458.

[4] Kiyotaki, N., and Wright, R.: A search theoretic approach to monetary economics, *American Economic Review* **83** (1993), 63–77.

[5] Kvasnička, M.: Markets, Social Networks, and Endogenous Preferences. In: *Proceedings of 30th International Conference Mathematical Methods in Economics* (J. Ramík and D. Stavárek, D., eds.), Silesian University, Karviná, 2012, 518–523.

[6] Moscarini, G., and Wright, R.: Introduction to search theory and applications. *Journal of Economic Theory*, **145** (2010), 1319–1324.

[7] Wilhite, A.: Economic activity on fixed networks. In: *Handbook of Computational Economics* (L. Tesfatsion and K. L. Judd, eds.), Elsevier, 2006, 1013–1045.

Examination of Internal Relations in Input-Output Factors of Competitiveness by SEM Method

Michaela Staníčková[1], Lukáš Melecký[2]

Abstract. Competitiveness has emerged as important discipline in area of strategy and research efforts have brought many interesting perspective and frameworks at national and regional level. Seeing competitiveness from competence point of view, role of factors lead to competitive success of countries and regions. The most important is the nature of these factors. Factors of competitiveness are internal and external, input or output type, endogenous or exogenous kind. The paper provides an analysis and evaluation of Regional Competitiveness Index approach (RCI). The main aim of the paper is an interception of internal links among factors of driven forces of competitiveness and internal links among factors of outcomes of competitive society and economy; and to model external relations between these two groups of input and output factors of competitiveness. By application of Structural Equation Modelling method is made simulation of own relationships between variables of RCI and it is thus investigated the importance of territorial characteristics for competitiveness within EU NUTS 2 regions in reference period 2004-2012.

Keywords: Competitiveness, input/output factors, NUTS 2, RCI, SEM method

JEL Classification: C15, C38, O18
AMS Classification: 62H20, 62H25, 93D25

1 Introduction

The term competitiveness is used in a bewildering variety of ways, both in the policy community and also in academic research. Policy makers in many countries frequently invoke competitiveness as a central objective of national economic policy, even though they often disagree about the ways to achieve it and there is also problem how to understand/how to define this concept. Many academic researchers, in contrast, have expressed scepticism about the term itself [6], partially due to some policies and practices of politicians that are put forward to promote competitiveness. A major problem is that there is no mainstream approach for competitiveness understanding – from this point of view, competitiveness is possible to define in different ways and levels, e.g. microeconomic, macroeconomic and regional competitiveness. There is a dichotomy in how policy makers think about competitiveness, but very often many definitions of competitiveness commonly used have at best an indirect connection to overall national economic performance [2]: on the one hand, competitiveness is associated with qualities that enable a high standard of living, on the other hand, competitiveness is associated with locational attributes that drive growth. Being an attractive location for investment affects prosperity indirectly and over the long run. These different views of competitiveness have confused the public and academic dialogue, and have obscured the development of an integrated framework to explain causes of cross-country differences in economic performance. The concept of competitiveness has quickly spread into the regional level, in the global economy regions are increasingly becoming the drivers of the economy and generally one of the most striking features of regional economies is the presence of clusters, or geographic concentrations of linked industries [9]. Current economic fundamentals are threatened by the shifting of production activities to places with better conditions. The regional competitiveness is also affected by the regionalization of public policy because of the shifting of decision-making and coordination of activities at the regional level. Within governmental circles, interest has grown in the regional foundations of national competitiveness, and with developing new forms of regionally based policy interventions to help improve the competitiveness of every region and major city, and hence the national economy as a whole. Regions play an increasingly important role in the economic development of states.

As very important for the materialization of local development and competitiveness policies is regarded the role of local authorities, especially with regards to the planning of and the implementation of local initiatives, promotion strategies and the city image on international level as well. This reflects the drive for growth mixed

[1] VŠB – TU Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, michaela.stanickova@vsb.cz.
[2] VŠB – TU Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, lukas.melecky@vsb.cz.

with the attempt to increase social responsibility, decrease worklessness and increase local entrepreneurship. The main priority developing these strategies focuses on the viable development of the local societies with an emphasis on more than just the economic development, since the existence of local authorities with entrepreneurial orientation derives as a basic need for future development. From this point of view, the paper investigates the importance of territorial characteristics on regional competitiveness within the European Union (EU) and the relations among driven forces of competitiveness and outcomes of competitive society and economy. Using Structural Equation Modelling (SEM) analysis, the importance of factors for the competitiveness of regions has been analysed.

2 Data and Structural Equation Modelling method

To improve the understanding of territorial competitiveness at the regional level, the European Commission has developed the Regional Competitiveness Index (RCI) which shows the strengths and weaknesses of each of the EU NUTS 2 regions. The RCI is based on eleven pillars describing both inputs and outputs of territorial competitiveness, grouped into three sets describing basic, efficiency and innovative factors of competitiveness; thus a large number of information regarding the most important determinants of competitiveness are included. Inputs describe driving forces of competitiveness, also in terms of long-term potentiality, and outputs are direct or indirect outcomes of a competitive society and economy [1]. With respect to this distribution, the RCI approach seems to be convenient as an initial dataset used in the paper. The RCI pillars consist of 66 indicators (RCI 2010) and 73 indicators (RCI 2013) which coming into creation of the RCI sub-indices. Because the RCI as a whole and its sub-indexes are latent variables which are indirectly represented by observed-manifest variables, it becomes important to select variables that are theoretically and intuitively appealing. Selection of observed variables, resp. indicators is based on territorial level which is subject of analysis – 272 NUTS 2 regions for 28 EU Member States. All RCI 2013 indicators are not used in the paper, because they were not available for each evaluated NUTS 2 region. In this paper, only 23 indicators are used – 6 for inputs and 17 for outputs. Input indicators represent (1) Motorway Transport - Length of Motorways (MTLM), (2) Air Transport of Freight (ATF), (3) Air Transport of Passengers (ATP), (4) Hospital Beds (HB), (5) Infant Mortality Rate (IMR), (6) Early Leavers from Education and Training (ELET). Output indicators represent (1) Employment Rate 15 to 64 years (ER), (2) Long-term Unemployment Rate (LtUR), (3) Unemployment Rate (UR), (4) Male Employment (ME), (5) Female Employment (FE), (6) Male Unemployment (ME), (7) Female Unemployment (FU), (8) Gross Domestic Product (GDP), (9) Compensation of Employees (CoE), (10) Employment in Sophisticated Sectors (EISS), (11) Human Resources in Science and Technology - Core (HRSTcore), (12) Patent applications to the EPO (EPO), (13) Total R&D Expenditure (GERD), (14) Human Resources in Science and Technology (HRST), (15) High-tech Patent Applications (HTI), (16) ICT Patent Applications (ICT), (17) Biotechnology Patent Applications (BioT). In the paper, indicators selected to populate the RCI 2013 framework are preferable of quantitative type (hard data) and mainly downloaded from Eurostat. Database is aggregated from reference years 2004 and 2012, which describe different period of economy cycles. Year 2004 characterize a growth period and year 2012 is possible to consider as post-crisis period.

For understanding the relations among input and outputs and their impact on competitiveness as a whole, the SEM analysis is applied. SEM can be viewed as a combination of factor analysis and regression or path analysis; it exhibits properties similar to multiple regression modelling, but more robustly takes into account relationships between (a group and sub-group of) latent variables, including correlations, covariances, nonlinearities, and error terms (correlated and uncorrelated). The basic idea is that, after the indirect and direct pathways that operate on the relationships of interest are defined, the latent variables, though they cannot observed by the researcher, can be estimated by their relation to observed variables (multiple indicators), see e.g. [7] or [10]. SEM thus allows for a set of observed variables to indirectly represent a composite latent variable in a relatively more objective fashion. A suggested approach to SEM analysis proceeds through the following process:

- literature review of the relevant theory and empirical research to support own model specification;
- specify a model (e.g., diagram, equations);
- determine model identification (e.g., the number of degrees of freedom for model testing is positive);
- select measures for the variables represented in the model;
- collect data;
- conduct preliminary descriptive statistical analysis (e.g., missing data, collinearity issues, outliers);
- estimate parameters in the model;
- assess model fit;
- respecify the model if meaningful and necessary;
- interpret and present results.

Regarding territorial and firms' competitiveness studies, SEM has been used in a variety of scientific works, see e.g. [8], [12] or [13]. Generally, SEM analysis is applied in order to simultaneously examine more complex relationships between observed and latent variables and to incorporate the latent variable of "Competitiveness" in the paper. A SEM model of the following form is utilized in order to examine the relationships of interest, namely the relationships of territorial competitiveness with factors that are established in the RCI as important indicators. It is common in general structural equation models such as the peer-influences model to distinguish between two sub-models [3]: 1. A structural sub-model relating endogenous to exogenous variables and to one-another (in the peer-influences model, the endogenous variables are unobserved, while the exogenous variables are directly observed); 2. A measurement sub- model relating latent variables (here only latent endogenous variables) to their indicators. The structural and measurement sub-models are written as follows:

$$\boldsymbol{\eta}_i = B\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\varsigma}_i, \tag{1}$$

$$\boldsymbol{y}_i = \boldsymbol{\Lambda}_y\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{2}$$

$$\boldsymbol{x}_i = \boldsymbol{\Lambda}_x\boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \tag{3}$$

$$i = 1,\ldots,N,$$

where N is number of observation, vector $\boldsymbol{x}_i$ represents observable exogenous variables (for observation *i*), $\boldsymbol{y}_i$ is vector of indicators of latent endogenous variables, $\boldsymbol{\eta}_i$ is vector of latent endogenous variables, $\boldsymbol{\xi}_i$ is vector of latent exogenous variables, $\boldsymbol{B}$ is matrix of structural parameters relating latent endogenous variables, $\boldsymbol{\Gamma}$ is matrix of structural parameters relating latent endogenous to exogenous variables, $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are factor loadings relating indicators to latent variables, $\boldsymbol{\varsigma}_i$ is vector of structural disturbances, $\boldsymbol{\varepsilon}_i$ are measurement errors in endogenous indicators and $\boldsymbol{\delta}_i$ is vector of of measurements errors in exogenous indicators.

The general structural equation model in following Fig. 1 includes observable (manifest) variables and unobservable endogenous (latent) variables. The observed variables compose of 23 indicators mentioned above (6 for inputs and 17 outputs); these indicators represent exogenous (external) variables influencing the factor of inputs and outputs (based on RCI approach) which have an impact on the level of Competitiveness. The latent variables are pillars in which are these indicators grouped, i.e. in the case of inputs (1) Infrastructure, (2) Health and (3) Education; in the case of outputs (4) Labour Market, (5) Market Size, (6) Business Sophistication and (7) Innovation. These latent variables indicating the latent factor of (8) Competitiveness and are symbolized η_i. Factor (1) – (8) are endogenous (internal) variables which are depended on initial indicators (exogenous variables). The method of estimation is Weighted Least Squares (WLS), analysing the matrix of polychoric correlations. The estimated parameters in the paper are standardized since they are considered more appropriate for a clearer interpretation and comparison of the estimated effects. Correlated error terms are introduced in the model as sawn in Figure 1. While correlated error terms are generally viewed with suspicion by researchers, however they can be justified in cases such as ours, where the observed indicators come from same database, reflecting very close meanings of competitiveness concept and have been introduced in a similar manner [7].

3 Findings of measuring the determinants of competitiveness by SEM

Results of the SEM model are presented in following Figure 1. Most of estimated factor loadings are positive and statistically significant as is possible to see from Figure1. Error terms are allowed to correlate as shown in Figure 1, all error term correlations are statistically significant. Regarding the latent factors (1) – (7) we can see that they are positively related to all of their indicators, i.e. manifest variables. In specific, "Infrastructure" is higher among regions who show higher level of air transport of freight and air transport of passengers, and there exist satisfactory range of motorway transport. Physical infrastructure clearly plays an important role in productivity. "Health" is also strongly related to its outcomes. Actually, the estimated correlation for the outcome variable infant mortality rate seems to be the highest among the observed estimated correlations within this latent variable. It is because of impact for the total number of population and future labour force. Indicator of hospital beds also shows higher level of correlation – the availability of hospital care is considered as a very important factor in the overall socio-economic situation in regions. The last latent variable in the case of input is "Education" and is represented by the only one indicator of early leavers from education and training. This indicator has problematic nature due to its relations with educated labour force and also with ability to succeed in the labour market, because if large parts of the population have limited basic reading and writing skills, their ability to actively participate in the economy is severely limited. These three latent variables are internal factors, i.e. factor endowment having direct influence on regional productivity and labour force mobilization for production the outputs.

First of output latent variable is “Labour Market” which consist of indicator about level of employment and unemployment; this variable has thus impact on labour availability, quality of local training and cultural factors that can affect the firms’ performance and productivity what has impact on regional competitiveness. Labour market policies are clearly important for the productivity of companies as well as for labour force mobilization. The strongest correlation is found for the outcome variable of female unemployment – working of women on the labour market and employment/unemployment is a long time issue. Effect of female labour is associated with a number of issues, such as changing the labour market orientation, population development, the number of men and women, resp. male and female labour force, etc. The lowest correlation is found for the outcome variable of unemployment rate – this indicator show rate of short-term, thus temporary unemployment and sessional unemployment. Much more important is to monitor the rate of long-term unemployment. “Market Size” is also related to its outcomes and the strongest correlation, as expected, is found for the outcome gross domestic product. GDP is the most important macroeconomic indicator represented the total economic activity of area and which is the indicator influenced by and influences all other indicators in the analysis. Also compensation of employees has positive correlation because refers basically to the total gross (pre-tax) wages paid by employers to employees for work done in an accounting period. The employment in sophisticated sectors is the only one indicator within the latent variable “Business Sophistication”. This indicator reflect the fact that modern employee relations are essential in order to compete within the new "knowledge economy” and also its correlation is corresponding. The last latent variable “Innovation” is, as expected, also strongly and positively related to its outcome indicators. Referred indicators take into account the innovation capability as a key driver for their advancement. These indicators refer to technological innovation which and to the technological readiness, which are dynamic factor of competitiveness. The strongest correlation are in the case of indicators human resources in science and technology, this reflects the fact that the bearer of progress and innovation is the man, resp. educated workforce.

When the relationship of the latent factors is examined, it is evident that “Competitiveness” is strongly and positively related to all seven variables, i.e. Infrastructure, Health, Education, Labour Market, Market Size, Business Sophistication and Innovation, indicating that these factors are important indicators for competitiveness of regions. The highest impact on regional competitiveness have quality and extent of infrastructure that have an impact on the accessibility of the region. Very high loadings are observed for the factors of education, labour market, innovation and health which are interlinked because the educated and healthy workforce is a key factor in increasing labour market participation and productivity and enhancing competitiveness at national and regional level. The innovation capability of the workforce is important because of the ability for improvement territorial productivity ‘simply’ adopting existing technologies or by invention innovative products and processes to maintain and improve productivity level. Also the correlation of market size and business sophistication is significant. These variables represent outputs of competitive society and economy and testify also about advancement of other latent factors which is manifested in indicators such as GDP, CoE and EiSS. The quantity and quality of workforce training, higher education and research in an economy all have a positive impact on prosperity as does the broader institutions and policies supporting innovation.

As expected, all the above mentioned input and output factors are significant indicators of regional competitiveness. “**Infrastructure**” exerting the higher effect among all input factors and “**Innovation**” has the highest correlation to competitiveness among all output factors. Based on scheme in Figure 1, all factors are shown to be highly associated with competitiveness since regional market characteristics, greater availability and quality of labour force and availability of adequate connections will facilitate production process, productivity and opening to other markets. Based on the fitness indices presented in Figure 1, it seems that the SEM model fits the data at acceptable level. Based on the findings and its comparison with the acceptable range, it can be stated that all fitness indices in the SBM model fall in the acceptable level. The hypothesis that the SBM model is correct is accepted, because Chi-square equals to **26.967** (measures the distance between data and hypothesis, the higher the value, the more distance of hypothesis from data) with **17** degrees of freedom (df) at probability level p = **0.435**. RMSEA (root mean square error of approximation) is **0.043** (a value of RMSEA of about 0.05 or less would indicate a close fit of the model in relation to the degrees of freedom). In Figure 1, there are the standardized estimates and squared multiple correlations displayed on the path diagram.

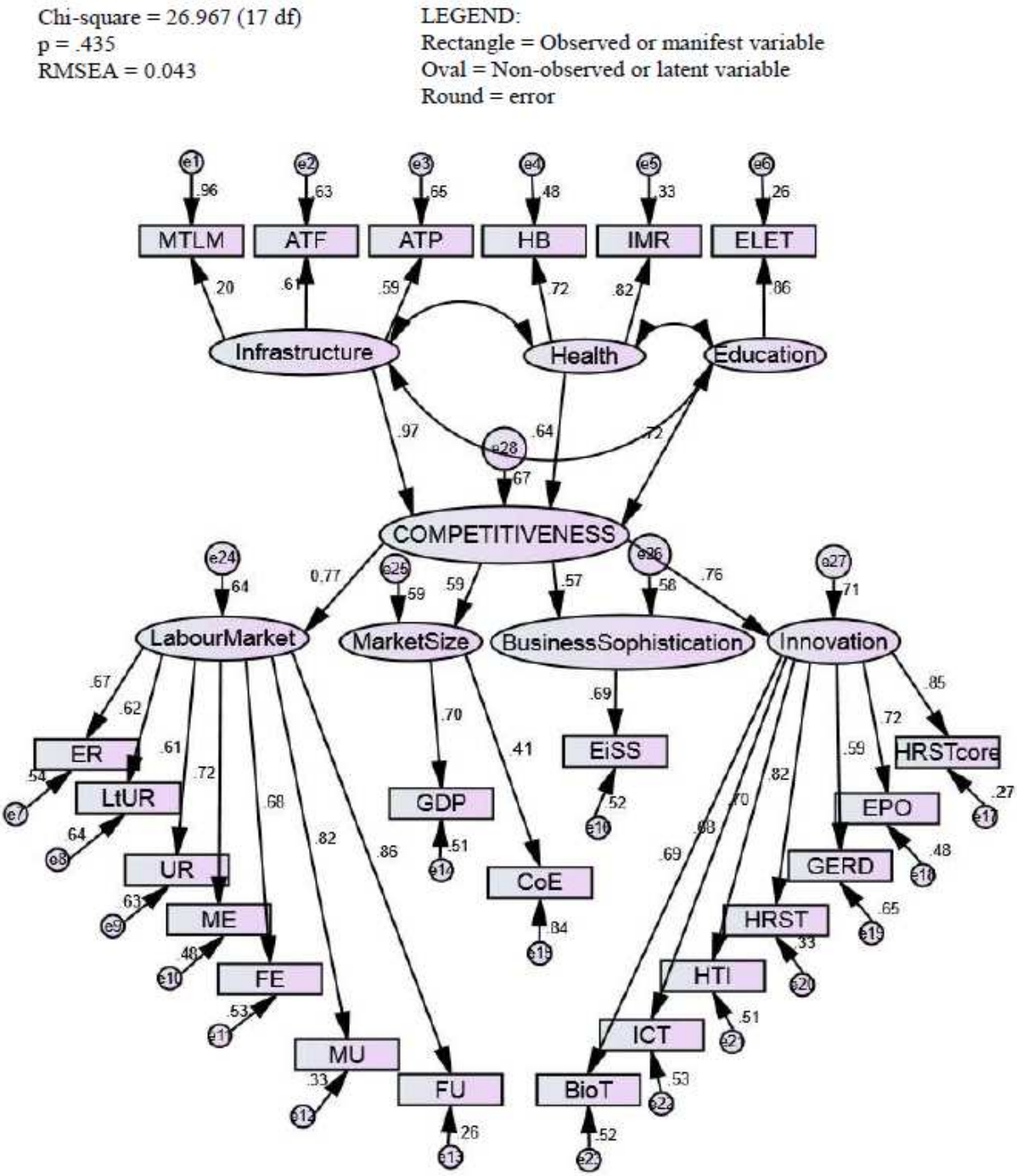


Figure 1 SEM Model for EU Regional Competitiveness

From the results is evident, that every territory has endowments, which are inherited and given, the prosperity is affected by these endowments directly. It is necessary to control for the effect of these endowments when estimating the role of competitiveness for economic performance. Regional competitiveness is driven by a range of institutions, policies, and public good investments that set the context for an entire economy. Physical and social infrastructure and local political institutions define the broader context in which productive economic activity takes place. Education, health care, and public safety are aspects of the overall social infrastructure necessary to enable productive economic activity. Factor conditions (quality and quantity) have long been seen as affecting competitiveness creation and productivity. The literature often makes a distinction between inputs (often from government investments) and incentives (competition, openness) as drivers of higher productivity. Company productivity is also strongly related to the set of incentives and rules that govern local competition and influenced thus level of regional competitiveness. High levels of competition on local markets are crucial for high performance. The vitality of competition affects the entry of new firms, the exit of underperforming old firms, and the performance patterns across existing firms – because firms operating in regions are mostly bearer

of outputs. But here are also other factor having impact on competitiveness. Prior studies have suggested geographic location, natural resource deposits, and country size are endowments that affect prosperity. A country's geographic location and its regions can affect the ease with which it can trade. Country size can attract foreign direct investments to access the local market and enable economies of scale in areas like R&D to be exploited [11]. Larger countries, resp. regions might more easily attract investments, even if they are not more competitive. However, empirical studies find that small countries enjoy greater openness than large countries or find no evidence of benefits of market size after controlling for the role of institutions [5].

4 Conclusion

The previous discussion leads to the conclusion that connectedness between aspects that participate in procedures of development and competitiveness at local level could contribute towards the direction of growth and progress. The most important issue is that the dynamic and ability of local authorities to plan and implement development policies and activities with the regional community so that the development procedure is guaranteed and controlled. Only through effective regional management and efficient use of sources to produce outputs it is possible to achieve a necessary level of competitiveness which will have a strong and healthy foundation. Nowadays, competitiveness is link with the ability to achieve certain overall outcomes, such as a high standard of living and economic growth. Competitiveness is also focused on the ability to achieve specific economic outcomes such as job creation, exports, or attraction of foreign direct investments. But competitiveness is also conditional by specific local conditions such as low wages, stable unit labour costs, balanced budget, or a 'competitive' exchange rate to support a current account surplus. The representation of all these common aspects is not simple but complicated and multidimensional. Regional development is thus not a random procedure but it is related with a number of factors the commonest of which is the structure of the local ability to attract direct foreign investments not only, but also on the ability to create the proper „business environment" where the firms will be able to operate effectively. But not only internal or external factor endowment is important for regional competitiveness, there is also important the significance of development partnerships between the public and private sector [4], there are advantages, as well as risks, in pursuing local economic development with involvement from the private sector. The promotion of sustainable development and the implementation of the partnership principle are two complementary, mutually reinforcing goals of EU policies aiming at successful local governance for enhancing regional competitiveness.

Acknowledgements

Supported by the SGS project (SP2014/111) of Faculty of Economics, VŠB-Technical University of Ostrava and the Operational Programme Education for Competitiveness – Project No. CZ.1.07/2.3.00/20.0296.

References

[1] Annoni, P., and Dijkstra, L.: *EU Regional Competitiveness Index 2013*. Publication Office of the European Union, Luxembourg, 2013.
[2] Delgado, M., Ketels, Ch., Porter, M.E., and Stern, S.: The Determinants of National Competitiveness. *Working Paper 18249*. National Bureau of Economic Research, Cambridge, 2012.
[3] Fox, J.: *An R and S-PLUS Companion to Applied Regression*. Sage Publications, Thousand Oaks, 2002.
[4] Getimis, P., and Grigoriadou, D. The Europeanization of Urban Governance in Greece: A Dynamic and Contradictory Process. *International Planning Studies*, 9 (2004), 5-25.
[5] Hall, R. E., and C. I. Jones.: Why Do Some Countries Produce So Much More Output per Worker than Others? *Quarterly Journal of Economics*, **114-1** (1999), 83-116.
[6] Krugman, P.: Competitiveness: A Dangerous Obsession. *Foreign Affairs*, **73-2** (1994), 28-44.
[7] Maruyama, G.M.: *Basics of Structural Equations Modelling*. SAGE Publications, California, 1997.
[8] Metaxas, T., and Economou, A.: Assessing the determinants of Firms' Competitiveness in Greece: A Structural Equation Modeling Analysis. *Working paper No. 42794*. Munich Personal RePEc Archive, Munich, 2012.
[9] Porter, M. E.: The Economic Performance of Regions. *Regional Studies*, **37-6/7** (2003), 549-578.
[10] Raykov, T., and. Marcoulides, G. A.: *A First Course in Structural Equation Modelling.* Lawrence Erlbaum Associates, Publishers, New Jersey, 2000.
[11] Romer, P. M.: Endogenous Technological Change. *Journal of Political Economy*, **98-5** (1990), 71-102.
[12] Santoso, E.B., Ananda,C.F., and Santoso, D.B.: Regional Competitiveness Analysis And Its Implication On Regional Development In East Java Region. In*: 23rd Pacific Conference of the Regional Science Association International (RSAI)*, Bandung, 2013, 1-23.
[13] Squalli, J., Wilson, K., and Hugo, S.: An Analysis of Growth Competitiveness. *Working paper No. 06-01*. Economic Policy Research Unit – Zayed University, Dubai, 2006.

On the use of genetic algorithms in the selection of predictors of parametric regression models.

Jacek Stelmach[1]

Abstract. Genetic algorithms deriving from the natural sciences have become very universal method also in statistical inference. The versatility arises from the possibility of various types of data encoding. The great strength of these algorithms is to reduce the efforts in searching for the optimal solution by using the analogy with the natural world - the processes of selection, crossover and mutation. In practice, it significantly reduces the time to choose such solution - from a huge number of permutations - which additionally are usually global optimum.
The experiment shows the results of the selection of explanatory variables in the parametric regression models, using genetic algorithms and comparing it with the most commonly used other methods. The experiment was performed for the empirical data related to the batch processes. In such processes, mostly dealing with huge amounts of potential explanatory variables, it is extremely difficult to select the right predictors - which do not increase unnecessary noise in the model.

Keywords: genetic algorithms, regression model, variable selection.

JEL Classification: C14
AMS Classification: 62H12

1 Introduction

One of the most important factor affecting the quality of the regression models is proper variable selection. Usually too many variables are measured – because either a researcher is not sure which one gives required information into a model or measurement system prepares a large amount of data. It, of course creates problems described in [5] and [7]:

1. a number of variables is irrelevant and disturbs a model as a noise;
2. a great number of variables can overfit the model and significantly increase a time of calculation;
3. significant correlation between variables creates numerical problems and add redundant information into the model.

Therefore there were developed the methods to select a subset of variables that should be a compromise between the complexity and goodness-of-fit of regression models.

The most obvious but very rarely used method is checking all the combinations of variables. But for bigger number of predictors it is almost impossible, so frequently the methods that make stepwise selection are used: forward selection, backward selection and stepwise multiple regression. The choice of the most proper subset bases on finding the optimum of fitness functions like: R^2, SE, AIC or BIC, see [7]. However dealing with the great number of variables can lead to local optimum which is not always a global. And as a result, regression model quality could be insufficient, see [8].. Therefore different heuristic methods are proposed to avoid such situation. One of them is Genetic Algorithm (GA).

2 Genetic Algorithms

Genetic Algorithms are class of the evolutionary algorithms invented by J. Holland in 1970s, see [6] to find the solutions to optimization and search problems – see [4]. It is the method that uses evolutionary approach inspired by the principles of biological evolution, coding a number of parameters into strings or chromosomes (genotypes) to generate another solution approaching the optimum. GAs select the best genotype to produce next generation. First, an initial population must be created. Let's assume that a genotype contains *k* binary genes like:

001011110101

The algorithm works faster when the initial population has got large variance that is maximized when 50% of genes are '0' and the others are '1'. The recommended methods is to randomly create a half of genotype and the other half create as a 'mirror' (changing '0' to '1' and vice versa), for example:

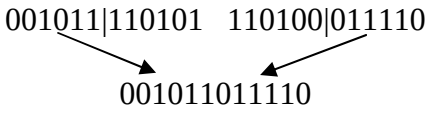
001101 010011

[1] University of Economics in Katowice, jacek.stelmach@polwax.pl

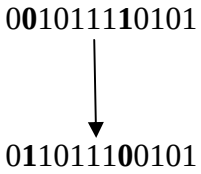
Subsequent generations are modified with the operations analogous to those found in biology (see figure 1.) described below, the computed fitness value for each population is used to:

- terminate the process (when a predefined number of generations was passed with no improvement of fitness value or a generations number reached specified value);
- make a selection among the genotypes.

1. selection – basing on chosen selection operator (fitness function), only genotypes with the best fitness value can be reproduced;
2. crossover – forms new offspring from two parent genotypes by combining part of the genes from each, randomly choosing the point (points) of crossover and the number of such points, for example:



3. mutation – it randomly alters the value of genes with certain probability, recommended (based on real evolution process) is to mutate few genes like:



Modifications of selection, crossover and mutation are presented wider in [3].

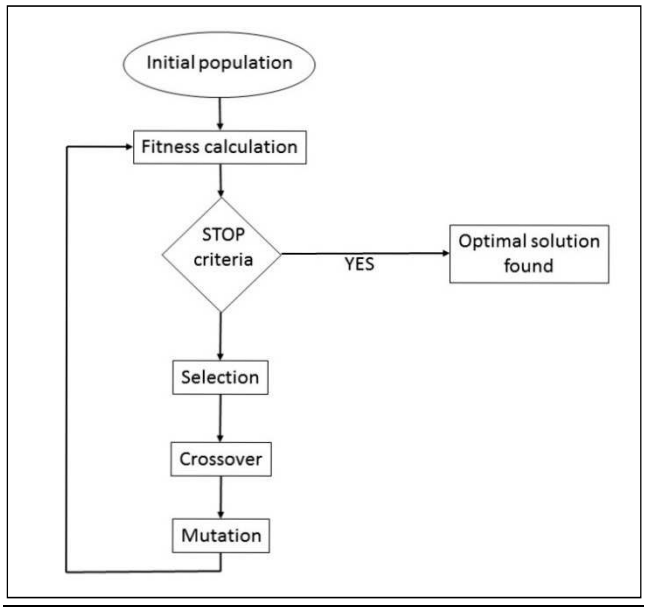


Figure 1 Sequence of genetic algorithm

3 Experiment description

3.1 Experiment design

The experiment was carried out for the one of the most known regression method: multiple linear regression minimizing the sum of squared error (OLS – Ordinary Least Squares). Taking into considerations a dataset containing n observations with k predictors: x_1, x_2, … x_k and one dependent variable y, the most obvious way is to create a genotype that codes a subset of explanatory variables in the following way:

- if i^{th} position is '0' – i^{th} predictor is not in the model';
- if i^{th} position is '1' – i^{th} predictor is included in the model'.

Such use of the GA is well-known. Presented experiment however expands the possibilities of the algorithm. Very often real processes and real phenomenon that we want to model are not linear. As a result, the quality of linear models is not sufficient enough and the predictions based on such models are inaccurate. Therefore GA used in this experiment determines not only the optimal subset of variables but also choose the most effective combination of the most popular non-linear functions: square, square root, logarithm and inverse. Therefore created genotype contains five times more genes than number of predictors and is divided into sections: original predictors, its squares, square roots, logarithms and inverses. This approach significantly increases the number of variables and although still stepwise method could be use, the probability of achieving a local optimum increases.

3.2 Experiment settings

In this study, an empirical dataset of twelve predictors with 108 observations and dependent variable ***average monthly price of hydrorefined paraffin* (*paraffin price)*** from the period: Jan 2003 – Dec 2011 was used. The predictors were chosen based on the experience for the data that specifies: raw material price, stock prices representatives of industries that use paraffin products and EUR to USD exchange rate with the delays for which the correlation between predictor and dependent variable were maximal. Proposed dependent variable is used in some cases as a parameter in price formulas. Therefore its model with good prediction capabilities can give a competitive advantage to companies using it:

1. X_1 - dependent variable delayed by 1 month;
2. X_2 - dependent variable delayed by 2 months;
3. X_3 - dependent variable delayed by 3 months;
4. X_4 – crude oil barrel price (USD) delayed by 2 months;
5. X_5 – crude oil barrel price (USD) delayed by 3 months;
6. X_6 – crude oil barrel price (USD) delayed by 4 months;
7. X_7 - EUR to USD exchange rate delayed by 5 months;
8. X_8 - EUR to USD exchange rate delayed by 6 months;
9. X_9 – Caterpillar Inc. stock price (USD) delayed by 2 months;
10. X_{10} - Goodyear Tire & Rubber Company stock price (USD) delayed by 6 months;
11. X_{11} - Freeport-MsMoran Copper&Gold Inc. stock price (USD) delayed by 3 months;
12. X_{12} - United States Steel Co stock price (USD) delayed by 3 months.

In order to better understand the properties of the proposed use of GA, the investigations were carried out additionally for well-known dataset '***forest fire***' that describes burned area of forest fire in the region of Portugal (http://archive.ics.uci.edu/ml/datasets/Forest+Fires) – 8 predictors with 517 observations and for three simulated dataset proposed by Friedman (see [1]) – 200 observations, according to formulas:

1. ***friedman 1***: $$y = 10\sin(\Pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + e_1 \quad (1)$$

2. ***friedman 2***: $$y = \sqrt{(x_1^2 + (x_2 x_3 - \frac{1}{x_2 x_4})^2)} + e_2 \quad (2)$$

3. ***friedman 3***: $$y = \tan^{-1}(\frac{x_2 x_3 - (x_2 x_4)^{-1}}{x_1}) + e_3 \quad (3)$$

where: x_1, x_2, x_3, x_4 – variables with uniform distribution of the intervals:

$0<x_1<100$; $40\Pi<x_2<560\Pi$; $0<x_3<1$; $1<x_4<11$;

e_1, e_3 ~N(0,1);

e_2 ~N(0,9)**.**

The AIC and BIC criteria have been considered as the possible fitness functions for GA (lower the value, the better the genotype), using linear-rank selection of populations with single point crossover and uniform random mutation. AIC (Akaike information criterion) and BIC (Bayesian information criterion) measure relative quality among regression models and are widely used for model selection purposes, see [5]. It was assumed that the calculations should stop if a number of created populations reached 1000 iterations or during 100 iterations there was no improvement of fitness value.

Presented above GA method was carried out to create optimal subsets for cases:

1. full set of variables including original data and four segments (root, square root, logarithm, inverse) – **AIC long, BIC long**;
2. the set limited only to original data – **AIC short, BIC short**.

Because one of the ways of dealing with too many variables and/or excessive correlation between predictors is PCA (Principal Component Analysis) transformation that converts a set of predictors into smaller number of new linearly uncorrelated variables, two additional cases were taken into considerations:

3. full set of variables after PCA transformation - **AIC long, BIC long**;
4. the set limited only to original data after PCA transformation – **AIC short, BIC short**.

The capabilities of multiple regression models created for optimal subset as above were compared using goodness-of-fit indexes: SE (standard error), R^2 (coefficient of determination), AIC, BIC with the subsets created with forward selection method for the same, described above dataset – **OLS long, OLS short**. The results presented in tables 1-5 show which variable was included into optimal model (black cell) and the values of goodness-of-fit index. For two last datasets (***friedman 3*** and ***forest fire***) Genetic Algorithm was not able to choose any variable for BIC fitness function. Notable is fast getting optimal solution - after only a few generations, see figure 2. An example describes AIC fitness value (precisely – the opposite value due to software limitations) calculated for full set of ***paraffin price*** dataset, depending on the generation number. But for all the experiments, the number of generations necessary to stabilize fitness value did not exceed 25.

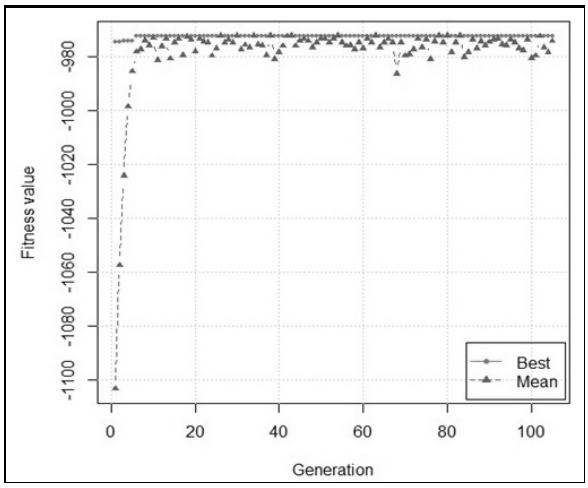


Figure 2 Reaching the optimal solution – an example

	Predictors	R^2	RSE	AIC	Var_no
OLS short		0,984	28,64	1020,8	6
AIC short		0,984	27,88	1020,3	6
BIC short		0,983	28,48	1020,8	4
OLS long		0,991	19,55	1001,5	44
AIC long		0,993	18,87	993,5	34
BIC long		0,991	20,95	1009,7	31
PCA transformation					
OLS short		0,984	27,83	1021,9	7
AIC short		0,984	27,83	1021,9	7
BIC short		0,984	27,83	1021,9	7
OLS long		0,993	21,29	998,2	55
AIC long		0,993	18,47	995,1	36
BIC long		0,989	23,50	1028,1	28

Table 1 Chosen variables and goodness-of-fit indexes for ***paraffin price*** dataset

	Predictors	R^2	RSE	AIC	Var_no
OLS short		0,647	2,683	972,3	4
AIC short		0,647	2,683	972,3	4
BIC short		0,647	2,683	972,3	4
OLS long		0,900	1,694	803,3	23
AIC long		0,884	1,536	799,3	29
BIC long		0,854	1,726	839,9	26
PCA transformation					
OLS short		0,655	2,653	977,9	9
AIC short		0,655	2,653	977,9	9
BIC short		0,646	2,688	981,0	8
OLS long		0,889	1,688	799,1	34
AIC long		0,883	1,546	795,9	31
BIC long		0,800	2,019	910,5	30

Table 2 Chosen variables and goodness-of-fit indexes for ***friedman 1*** dataset

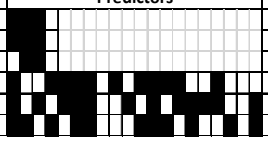
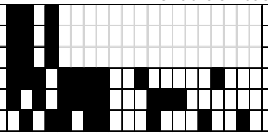
	Predictors	R^2	RSE	AIC	Var_no
OLS short		0,885	129,727	2521,7	3
AIC short		0,885	129,727	2521,7	3
BIC short		0,883	131,076	2523,9	2
OLS long		0,893	126,200	2521,2	11
AIC long		0,895	124,165	2522,2	12
BIC long		0,894	124,524	2521,4	11
PCA transformation					
OLS short		0,885	129,694	2521,6	4
AIC short		0,885	129,694	2521,6	4
BIC short		0,885	129,694	2521,6	4
OLS long		0,895	125,700	2517,8	9
AIC long		0,894	124,374	2516,9	9
BIC long		0,888	127,980	2526,3	8

Table 3 Chosen variables and goodness-of-fit indexes for ***friedman 2*** dataset

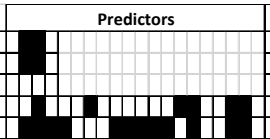
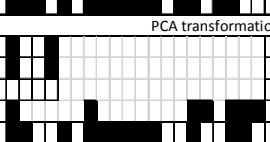
	Predictors	R^2	RSE	AIC	Var_no
OLS short		0,040	0,998	572,7	2
AIC short		0,040	0,998	572,7	2
BIC short					
OLS long		0,071	0,994	600,3	6
AIC long		0,083	0,975	583,5	12
BIC long		0,083	0,975	587,5	14
PCA transformation					
OLS short		0,040	0,997	572,5	2
AIC short		0,040	0,997	572,5	2
BIC short					
OLS long		0,075	0,995	604,9	7
AIC long		0,086	0,973	582,8	12
BIC long		0,067	0,984	585,0	11

Table 4 Chosen variables and goodness-of-fit indexes for ***friedman 3*** dataset

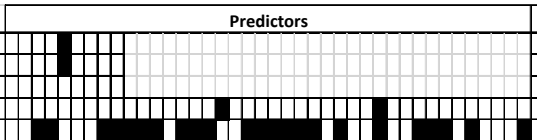
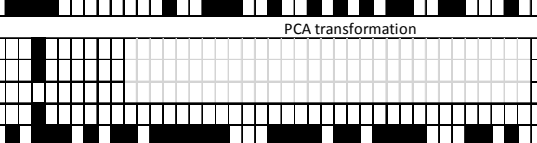
	Predictors	R^2	RSE	AIC	Var_no
OLS short		0,010	63,412	5761,9	1
AIC short		0,010	63,412	5761,9	1
BIC short					
OLS long		0,012	64,390	5799,2	2
AIC long		0,023	62,975	5798,8	28
BIC long		0,020	63,066	5788,3	17
PCA transformation					
OLS short		0,006866	63,49847	5763,329	1
AIC short		0,006866	63,49847	5763,329	1
BIC short					
OLS long		0,006866	63,49847	5763,329	1
AIC long		0,025031	62,91506	5803,785	31
BIC long		0,020625	63,05707	5796,116	21

Table 5 Chosen variables and goodness-of-fit indexes for ***forest fire*** dataset

The results of the investigations show that the model based on a subset of variables chosen with genetic algorithm is better fit, even with the smaller number of predictors than chosen with forward selection method. It is especially visible for ***paraffin price*** and ***friedman 1*** datasets. Analyzing goodness of fit indexes, better results are achieved using AIC criteria as a fitness function. Additional benefit is selecting the predictors from a set of non-linear functions of original variables what confirms the non-linear nature of the real phenomena. Very interesting is similar results comparing subset from original dataset and from dataset after PCA transformation although independent on selection method, there were chosen also predictors correspond to smaller eigenvalues. PCA transformation gives, however additional benefit - the ability to create models even if the predictors are highly correlated. Unfortunately, ***forest fire*** and ***friedman 3*** datasets are not the source of sufficient data to create a good regression models, regardless of the method of selection of variables, what only confirms that there is no universal method, if the quality of the obtained observation is inadequate.

4 Conclusions

Presented results indicate that the proposed algorithm allows to create the regression models with a higher goodness-of-fit, although the number of selected variables is usually less than the number of variables selected with forward selection. Clearly visible is also a better fit of models using additionally a non-linear functions of the predictors. Better results (recommended) is to use as a fitness function – AIC criteria. The same behavior is observed for the predictors after PCA transformation, however comparing the dataset with and without PCA transformation, goodness-of-fit the best subsets for AIC fitness function is comparable. The proposed method does not allow for a good fit, of course, for all the models. For two last cases, for which it is hard to create good models, even with the non-parametric methods - the best selected subsets do not guarantee a good fit, regardless of the methods used for selection of predictors purposes.

Acknowledgements

The work paper funded by resource on science period 2 0012-2015, National Science Centre Nr DEC-2011/03/B/HS4/05630.

References

[1] Friedman, J.: Multivariate adaptive regression splines. *Annals of Statistics* **19**(1991), 37-44.
[2] Giustolisi, O.: Using a multi-objective genetic algorithm for SVM construction, *Journal of Hydroinformatics* **8.2** (2006), 125-139.
[3] Grzymkowski, R., Kaczmarek, K., Kiełtyka, S., Nowak, I.: *Wybrane algorytmy optymalizacji. Algorytmy genetyczne. Algorytmy mrówkowe*, Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, Gliwice 2008.
[4] Goldberg, D.: *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Boston, 1989.
[5] Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*, Springer, 2008.
[6] Holland, J..: *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, M1, USA, 1975.
[7] Rencher, A. C.: *Methods of Multivariate Analysis*, A John Wiley & Sons Inc., 2002.
[8] Timm, N. H.: *Applied Multivariate Analysis*, Springer, 2002.
[9] Zhang, B., Horvath, S.: Ridge regression based hybrid genetic algorithms for multi-locus quantitative trait mapping, *International Journal of Bioinformatics Research and Application* **1.3**(2006), 261-272.

Classifier performance assessment in social science - does the quality of data matter?

Jan Stoklasa [1], Pasi Luukka [2], Jana Talašová [3]

Abstract. In social science it is common to deal with data of variable quality, as a direct consequence of having human beings as the main source of information. Various measures of data quality have been developed in humanities e.g. validation scales or lie scores. In many real life cases, decisions need to be made or diagnoses assigned based on data of lower quality when data of higher quality are not available - any information on its quality is therefore valuable. Real life decision making problems in social science can be translated into the language of classification. In mathematical classifier design theory various measures of classifier performance have been designed. As far as we know, none of these reflects directly the quality (reliability) of the data. On the example of the receiver operating characteristics (ROC) and the area under the ROC curve (AUC) introduced in signal detection theory we show, how data quality can be incorporated into performance assessment of classifiers. We present a modification of the ROC approach reflecting data quality and discuss possible benefits of its use on artificial data.

Keywords: classification, ROC, ROC curve, area under curve, AUC, quality.

JEL classification: C44
AMS classification: 90B50, 62C86

1 Introduction

The study of human beings, their behavior, decision making, environment, interactions and so on is in the focus of humanities. As human beings are in the center of attention, they are also a crucial source of data. Methods of gathering data in humanities range from measuring certain attributes through observation to interviews with people. Human beings, who provide much of the data, can however influence the quality (information value) of the data. People can try to conceal certain pieces of information, try to look better, lie and so on. These distortions of data can not be completely eliminated and can significantly influence the results of the decision making process based on them. Hence in the areas of humanities, where people themselves and their introspection and self report (that is not objective measurement) are the main source of data, researchers and above all practitioners need to be able to deal appropriately with data distortion.

The first step is the identification of data distortion and its quantification (however fuzzy, imprecise or uncertain this might be). The second natural step is reflecting this information in the decision making processes. From the mathematical point of view, this means reflecting data quality obtained through already existing measures (or designing appropriate measures of data quality) in the mathematical models used in humanities. In this paper, we take the diagnostics processes typical for medicine or psychology as an example of such situations, where distortions of data may exists and conclusions still need to be made. These situations can be seen as classification tasks, where each instance (a person represented by a collection of data) needs to be assigned a class (diagnosis) that is most appropriate based on the characterising data.

[1]Palacký University in Olomouc/Faculty of Science, Department of Mathemetical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46, Olomouc, Czech Republic, jan.stoklasa@upol.cz

[2]Lappeenranta University of Technology/School of Business and Deptartment of Mathematics and Physics, Skinnarilankatu 34, 53850 Lappeenranta, Finland, pasi.luukka@lut.fi

[3]Palacký University in Olomouc/Faculty of Science, Department of Mathemetical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46, Olomouc, Czech Republic, jana.talasova@upol.cz

Our aim is to discuss classification under such circumstances, when the validation set contains instances of variable data quality. Particularly, we aim to discuss here the issue of classifier performance assessment under variable data quality. We illustrate our point on a practical example from psychological diagnostics and propose a possible solution to this problem by modifying one of the widely used classifier performance measures - the receiver operating characteristics (ROC) curve and the area under this curve (AUC). The paper is organised as follows. In the next section, short summary of necessary mathematical concepts is provided. Section 3 provides a short insight into data quality assessment in psychological diagnostics (data validity in MMPI-2 is presented as an example). In Section 4 we specify the problem and give some rationale for the modification of classifier performance assessment methods presented in Section 5. We discuss classic ROC analysis and our proposed modification of this approach in Section 5 and in Section 6 a numerical example and comparison of the classic and modified ROC approach is presented. Discussion and conclusions follow.

2 Preliminaries

In the next section, we will provide an example of a fuzzy set based approach to data quality assessment through certain so called *validity scales* used in the MMPI-2 (a personality inventory used in psychological diagnostics, see e.g. [5]). We will therefore briefly recall the notion of fuzzy sets here as introduced by Zadeh in [11] or further elaborated in [1].

Let U be a nonempty set (the universe of discourse). A *fuzzy set* A on U is defined by the mapping $A : U \to [0,1]$. For each $x \in U$ the value $A(x)$ is called a *membership degree* of the element x in the fuzzy set A and $A(.)$ is called a *membership function* of the fuzzy set A. The symbol $F(U)$ denotes the family of all fuzzy sets on U. The *height of* a fuzzy set A is a real number $\mathrm{hgt}(A) = \sup_{x\in U}\{A(x)\}$. If the fuzzy set A represents the meaning of a linguistic term (say a linguistic description of a property P of the elements of the universe U, e.g. *"tall"* or *"valid"*), the membership degree of each element $x \in U$ in the fuzzy set A, $A(x)$, can be interpreted as the compatibility of x with the linguistic description of the property (how well is x described by P). In the case when a fuzzy set A on U is used to represent a meaning of a linguistic term we require the fuzzy set A to be normal, that is $\mathrm{hgt}(A) = 1$. This means that there exists at least one such $x \in U$ that is fully compatible with P.

Other important concepts related to fuzzy sets are: a) the *kernel* of A, $\mathrm{Ker}(A) = \{x \in U \mid A(x) = 1\}$; b) the *support* of A, $\mathrm{Supp}(A) = \{x \in U \mid A(x) > 0\}$; and c) for $\alpha \in [0,1]$ the α-*cut* of A, $[A]_\alpha = \{x \in U \mid A(x) \geq \alpha\}$. If the support of A is a finite set, $\mathrm{Supp}(A) = \{x_1, \ldots, x_k\}$, then the fuzzy set A will be described as $A = \{{}^{A(x_1)}\!/_{x_1}, \ldots, {}^{A(x_k)}\!/_{x_k}\}$. The cardinality of such a fuzzy set A is given by $Card(A) = \sum_{i=1}^{k} A(x_i)$. Any crisp set $\{x_1, \ldots, x_k\}$ can be represented by the fuzzy set $\{{}^{1}\!/_{x_1}, \ldots, {}^{1}\!/_{x_k}\}$.

A *union* of fuzzy sets A and B on U is a fuzzy set $A \cup B$ on U with the membership function for all $x \in U$ given by $(A \cup B)(x) = \max\{A(x), B(x)\}$. An *intersection* of fuzzy sets A and B on U is a fuzzy set $A \cap B$ on U with the membership function for all $x \in U$ given by $(A \cap B)(x) = \min\{A(x), B(x)\}$. Let A be a fuzzy set on U and B be a fuzzy set on V. Then the Cartesian product of A and B is the fuzzy set $A \times B$ on $U \times V$ with the membership function defined for all $(x, y) \in U \times V$ by $(A \times B)(x, y) = \min\{A(x), B(y)\}$.

We will also recall here a very general definition of a classifier. Let $\mathbf{x}$ be a vector in an n-dimensional real space $\mathbb{R}^n$ and let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ be a set of class labels. A (crisp) *classifier* is any mapping $D : \mathbb{R}^n \to \Omega$. If Ω were a closed interval, we would call D a *continuous classifier*. This general definition will be enough for the purpose of this paper. Continuous classifiers, where $\Omega = [0, 1]$, 0 and 1 represent the two limit states (binary classifier) will be considered in this paper. This way the diagnostics process in medicine or psychology can be seen as a classification problem with the following two classes: a person is healthy (corresponds with 0) or not healthy (corresponds with 1). The fact that the output of a continuous classifier may lie between the points 0 and 1 reflects that a person can only partially meet the criteria for "healthy" or "not healthy". A binary classification can be achieved by introducing a threshold $t \in [0, 1]$ such that state 0 is assigned when the classifier output is lower than t, and state 1 otherwise.

3 Possible measures of data quality - an example from psychology

Before we proceed to the definition of the problem, let us briefly illustrate how the quality of data can be captured and mathematically represented. We will do so on an example of psychological diagnostics

method - the Minnesota Multiphasic Personality Inventory (MMPI-2). MMPI (and its revised version MMPI-2) are test methods for psychological differential diagnostics developed by Hathaway and McKinley in 1930s (first introduced in [6]). The MMPI-2 version of the test consists of 567 statements to which the tested person provides "agree/disagree" answers. The answers saturate 10 basic clinical scales, 7 basic validity scales, many content scales and supplementary scales (summary of these values is called a *protocol*). Validity is seen here as a multifactorial phenomenon that can be assessed through various criteria. These criteria reflect for example the number of unanswered items, consistency of answers to similar items, desire to appear better or worse, bizarre answers or effort to hide or stress pathological symptoms. A valid protocol is such that its validity is acceptable according to all the validity criteria. For each criterion a fuzzy set describing its "acceptable values" can be defined (using the respective computed scores as a universe, see [10] for more details; Figure 1 gives an example for one validity criterion). We can define a prototype of a valid MMPI-2 protocol in the following way

$$VP = M(VC^1) \times \cdots \times M(VC^7), \tag{1}$$

where $M(VC^1), \ldots M(VC^7)$ are the fuzzy sets representing the meanings of *acceptable values* of each validity criterion. Let us consider a set of n protocols. If we denote $vc^1, \ldots, vc^7$ the crisp values of each validity criterion for a given protocol, we can define the quality of each protocol i (that is its validity rate) as $q_i = VP(vc_i^1, \ldots, vc_i^7)$, for all $i = 1, \ldots, n$. This way we have defined the quality of each protocol (data instance) as the membership degree to the prototype of valid protocol, obviously $q_i \in [0, 1]$.

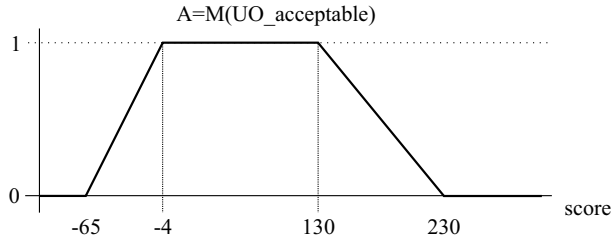


Figure 1 Fuzzy number representing the meaning of an "acceptable score" for the UO scale.

4 Quality of data and classification - problem specification

It is not our aim in this paper to discuss various classifiers, their principles, advantages and possible shortcomings. Regardless of what mathematical theory a classifier is based on, at the end of the day we need to have a means of assessing how well it performs on a given data set. And this assessment has to be done as reasonably as possible. We will assume here, that the classifier (any mapping, that assigns class labels to members of some universe of discourse) is trained on some data set. This training data set can either be an actual data set, or in the case of expert knowledge based classifiers, the training data set can be implicitly present in the knowledge of the expert. It is important to realize, that we usually assume that the data on which we train our classifiers (or on which our expert gained his/her know-how that is then represented by the classifier) are of high quality. In fact the assumption is that the data in the training set is either of high quality, or if high quality is unachievable, at least the data instances are of comparable quality which is as high as we are able to assure. We usually do not consider training a classifier on data of poor quality.

As a result of this assumption, we obtain a classifier supposedly working well on quality data after its training phase is completed. The question we now have to face when we apply the classifier in the field of humanities is what if not all the data is of high quality? What if the quality of data varies in the validation set or in real life? Our particular interest in this paper is to propose how any classifier performance assessment tool should reflect the quality of data in the validation set. For simplicity's sake let us consider binary classifiers. A performance measure of a binary classifier has to take into account the number (or at least the ratio) of well classified objects and the number (ratio) of misclassified objects - see Section 5 for an example of such performance measure - the AUC. This is quite straightforward when the quality of all data instances is high or at least similar.

In case of big differences in the quality of data instances, we might want to reflect the quality of data as well. The rationale for this claim is simple - the conclusions we can draw when the classifier

makes a misclassification on a high quality data instance and on a low quality one are quite different. Misclassification of an instance of data of low quality is understandable, as the classifier was trained on high quality data instances and low quality of data means lower information value (and problematic interpretability). Misclassifications of high quality data instances are on the other hand intolerable. As such misclassification of a high quality data instance should lower the performance indicator for the classifier much more than a misclassification of a low quality data instance. Analogically for correct classifications.

5 Measuring classifier performance - standard and modified ROC

Let us now consider method for binary classifier performance assessment widely used in medicine - the Receiver Operating Characteristics (ROC) analysis. This method and its origin can be traced to signal detection theory (SDT) and it was further extended in the psychological context by [4] to graphically represent the performance of classifiers (see [2] for more details). As we do not have enough room to discuss the background of SDT here, we direct the readers to e.g. [2, 3, 9] for its more detailed description and fuzzy extension. The SDT was developed to assess the performance of binary classifiers. Two outputs of the classifier are distinguished - *positive* (P - identified here with numerical value 1 and label "*not healthy*") and *negative* (N - identified here with numerical value 0 and label "*healthy*"). For continuous classifiers, a threshold $t \in [0, 1]$ is set to assess the performance of the classifier. Four sets of instances are then distinguished for a given t (if classifier output is lower than t then N is assigned, otherwise P):

TP_t (true positive) - set of instances where *not healthy* were classified as *positive* (P),

FP_t (false positive) - set of instances where *healthy* were classified as *positive* (P),

FN_t (false negatives) - set of instances where *not healthy* were classified as *negative* (N),

TN_t (true negatives) - set of instances where *healthy* were classified as *negative* (N).

Let us denote now $H \subseteq \{x_1, \ldots, x_n\}$ the set of healthy instances and $NH \subseteq \{x_1, \ldots, x_n\}$ be a set of not healthy instances, H and NH are disjunctive and their union is the set of all instances. We can now compute two important characteristics *sensitivity* (TP_rate_t) and *specificity* ($1 - FP\_rate_t$).

$$TP\_rate_t = \frac{Card(TP_t)}{Card(NH)} = \frac{Card(TP_t)}{Card(TP_t \cup FN_t)} \tag{2}$$

$$FP\_rate_t = \frac{Card(FP_t)}{Card(H)} = \frac{Card(FP_t)}{Card(TN_t \cup FP_t)} \tag{3}$$

Here $Card(X)$ denotes the cardinality of a set X, i.e. the number of its elements. Now an $ROC\_curve = \{(x, y) \mid x = FP\_rate_t, y = TP\_rate_t, \text{for all } t \in [0, 1]\}$ can be plotted. For continuous classifiers the area under the ROC curve (AUC) can be computed. The AUC serves as a measure of the performance of the classifier (an ideal classifier has $AUC = 1$; classifiers with $AUC > 0.5$ are considered better than random classification - see [3, 7] for more details). The AUC can be interpreted as the probability that a randomly chosen *not healthy* instance will be assigned greater value by the classifier, than a randomly chosen *healthy* instance. It is however obvious, that this approach to ROC does not reflect the quality of data. We will discuss this issue on a numerical example after we propose a modification of this approach.

To reflect the quality of the data in the spirit of Section 4, let us make the following modifications to the ROC analysis. We will make use of the quality measure $q_i \in [0, 1]$ - *quality rate* of a data instance i. This way each instance or object $x_i \in \{x_1, \ldots, x_n\}$ we want to classify can be assigned a quality rate q_i. We can define a fuzzy set of quality data instances as $\widetilde{QDI} = \{{}^{q_1}\!/_{x_1}, \ldots, {}^{q_n}\!/_{x_n}\}$. This fuzzy set will play an important role in the proposed modification of the classic ROC analysis. Let us suppose, that the continuous classifier the performance of which we are to assess assigns to each instance $x_i \in \{x_1, \ldots, x_n\}$ a value $cl_i \in [0, 1]$. We will now define the fuzzy sets of not healthy, healthy, true positive, false positive, false negative and true negative instances respectively for any value $t \in [0, 1]$ and for all $i = 1, \ldots, n$:

$\widetilde{NH} = \{{}^{s_1}\!/_{x_1}, \ldots, {}^{s_n}\!/_{x_n}\}$ and $\widetilde{H} = \{{}^{1-s_1}\!/_{x_1}, \ldots, {}^{1-s_n}\!/_{x_n}\}$, where $s_i = 1$ iff x_i is *not healthy* and $s_i = 0$ otherwise,

$\widetilde{TP_t} = \{{}^{\alpha_1}\!/_{x_1}, \ldots, {}^{\alpha_n}\!/_{x_n}\}$, where $\alpha_i = 1$ iff $[(cl_i \geqq t) \wedge (s_i = 1)]$ and $\alpha_i = 0$ otherwise,

$\widetilde{FP_t} = \{^{\beta_1}/_{x_1}, \ldots, ^{\beta_n}/_{x_n}\}$, where $\beta_i = 1$ iff $[(cl_i \geqq t) \wedge (s_i = 0)]$ and $\beta_i = 0$ otherwise,

$\widetilde{FN_t} = \{^{\gamma_1}/_{x_1}, \ldots, ^{\gamma_n}/_{x_n}\}$, where $\gamma_i = 1$ iff $[(cl_i < t) \wedge (s_i = 1)]$ and $\gamma_i = 0$ otherwise,

$\widetilde{TN_t} = \{^{\delta_1}/_{x_1}, \ldots, ^{\delta_n}/_{x_n}\}$, where $\delta_i = 1$ iff $[(cl_i < t) \wedge (s_i = 0)]$ and $\delta_i = 0$ otherwise.

These are in fact crisp sets represented as fuzzy sets to be compatible with $\widetilde{QDI}$ in the following computations. Now we can define the modified characteristics of the classifier using the above mentioned definitions and $\widetilde{QDI}$. For any value of threshold $t \in [0, 1]$ we now propose to include the information regarding the quality of data in computing the following characteristics:

$$TP\_rate_t = \frac{Card(\widetilde{TP_t} \cap \widetilde{QDI})}{Card(\widetilde{NH} \cap \widetilde{QDI})} = \frac{Card(\widetilde{TP_t} \cap \widetilde{QDI})}{Card[(\widetilde{TP_t} \cup \widetilde{FN_t}) \cap \widetilde{QDI}]}, \tag{4}$$

$$FP\_rate_t = \frac{Card(\widetilde{FP_t} \cap \widetilde{QDI})}{Card(\widetilde{H} \cap \widetilde{QDI})} = \frac{Card(\widetilde{FP_t} \cap \widetilde{QDI})}{Card[(\widetilde{TN_t} \cup \widetilde{FP_t}) \cap \widetilde{QDI}]}. \tag{5}$$

We can now construct the ROC curve analogically as in the crisp case and compute the AUC.

6 Numerical example

To see how the classic ROC performs on data compared to the modified version that takes into account the quality of data, 10 000 data instances were generated - for each instance a classifier output, quality rate and a proper class was generated (sampled from uniform distribution). Quality rate was then modified for all the instances in four variants: a) no quality rate used, b) instances classified correctly were assigned higher quality rate [1], c) misclassified instances were assigned higher quality rate[2], d) quality remained randomly assigned. The AUC was then computed for all four variants. Figure 2 summarizes the results.

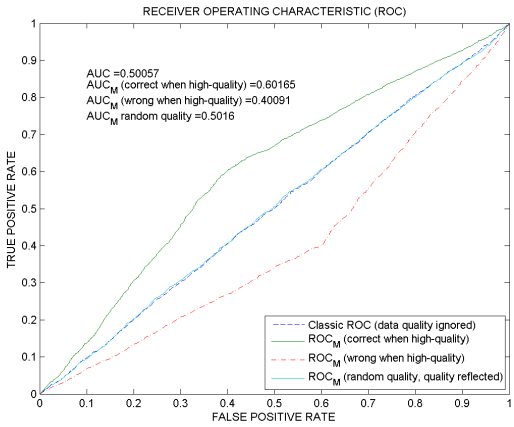


Figure 2 ROC curves and the AUC computed for 10 000 instances of artificial data in four variants: a) Classic ROC - quality of data ignored (near diagonal, dashed line), b) Modified ROC (correct on high-quality data; solid green), c) Modified ROC (wrong on high-quality data; dashed red), d) Modified ROC (random quality; near diagonal, solid).

7 Discussion

Results summarized in Figure 2 show, that when there is no pattern in the mistakes - that is misclassifications are distributed randomly regardless of data quality, the modified (case d) and the classic (case

[1] Correct classification in this case was considered to mean $|cl_i - s_i| < 0.5$. If this condition was met then $q_i = \min\{1, 0.5 + (0.5 * rand_{[0,1]} - 0.3 * rand_{[0,1]})\}$, otherwise $q_i = \max\{0, 0.5 - (0.5 * rand_{[0,1]} - 0.3 * rand_{[0,1]})\}$, where $rand_{[0,1]}$ is a random number generated from $[0, 1]$ under uniform distribution.

[2] Analogically to case b).

a) ROC provide both almost identical values of AUC around 0.5, as was expected. If the classifier tends to make mistakes on high quality data, its performance is assessed as much worse (case c) - AUC drops to about 0.4). On the other hand if the misclassifications are associated with lower-quality data, the performance of the classifier is assessed as much better (case b) - AUC rises to about 0.6). The modification of classic ROC described here makes a standard classifier performance measure (AUC) able to reflect the quality of the data. The result of the proposed modification is, that data with higher quality influence the value of the performance measure more than low quality data. Data with zero quality have no direct effect on the performance measure represented here by the AUC. ROC and AUC serve here as an example of a classifier performance measure - more measures are used in practice, but as far as we know none of them provides explicit means for including the quality of data from the validation set into the performance assessment process. The important idea resulting from our research (and supported by the results of the simulation) is that when the quality of data can be assessed, this information should be taken into account when evaluating classifier performance. It seems reasonable to "tolerate" mistakes on low quality data and require precision on high quality data instances. This is even more necessary when data quality is really different among the instances and can not be improved.

8 Conclusions

In this paper, we have proposed a useful modification of a classifier performance assessment technique frequently used in medicine and psychology - ROC. This modification makes it possible to reflect the quality of the data. This way the performance of the classifier represented here by the AUC really reflects the classification success rate while tolerating mistakes on low quality data. This allows for a more objective assessment of classifier performance.

Acknowledgements

This research was supported by the grant *GA 14-02424S* of the Grant Agency of the Czech Republic.

References

[1] Dubois, D. & Prade, H.: *Fundamentals of fuzzy sets.* Kluwer Academic Publishers, Dordrecht, 2000.

[2] Egan, J.: *Signal detection theory and ROC analysis.* Academic, New York, 1975.

[3] Fawcett, T.: *ROC graphs: notes and practical considerations for researchers.* Technical report HPL-2003-4, HP Labs, 2004.

[4] Green, D. M. & Swets, J. A.: *Signal detection theory and psychophysics.* Wiley, New York, 1966.

[5] Greene, R. L.: *The MMPI-2: An interpretive manual.* Allyn and Bacon, Boston, 2000.

[6] S. R. Hathaway and J. C McKinley: A multiphasic personality schedule (Minnesota): I. Construction of the schedule. Journal of psychology, **10** (1940), 249-254.

[7] Krzanowski, W. J. & Hand, D. J.: *ROC curves for continuous data.* London: Chapman and Hall, 2009.

[8] Kuncheva, L. I.: *Fuzzy classifier design.* Physica Verlag, Heidelberg; New York, 2000.

[9] Parasuraman, R., Masalonis, A. J. & Hancock, P. A.: Fuzzy signal detection theory: basic postulates and formulas for analyzing human and machine performance. *Human factors,* **42**(4) (2000), 636-659.

[10] Stoklasa, J. & Talašová, J.: Using linguistic fuzzy modeling for MMPI-2 data interpretation. In *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011 - part II, 2011, Praha, Czech Republic*, 2011, 653-658.

[11] Zadeh, L. A.: Fuzzy sets. *Inform. Control*, 8(3) (1965), 338-353.

The influence of world oil prices on wheat price level

Radmila Stoklasová[1]

Abstract. Central banks are currently focused on the role of asset prices in monetary policy and financial stability decision making. The attention of most small open economies is focused on the development of the financial markets (stock prices, bond prices and exchange rates), but also on commodity prices, particularly oil. Oil plays an important role in the global economic system. This paper deals with dependencies between the price of oil and the price of wheat. There is investigated short-term and long-term relationship between the commodities using time series analysis. There are used methods of correlation and regression analysis, Granger causality, Augmented Dickey-Fuller (ADF) test of stationarity. For time series analysis (monthly average commodity prices, October 2004 – February 2014) there were used computer programs SPSS and GRETL.

Keywords: Time series analysis, correlation analysis, regression analysis, Granger causality, ADF test of stationarity.

JEL Classification: C13, C22
AMS Classification: 62-07, 62J05

1 Basic characteristics

The prices of commodities are quoted in U.S. Dollars (USD) for 1 bushel for wheat and for oil it is for 1 barrel. In [2], [4], [5] we can see more information. The analysis was based on data from May 2004 to February 2014, while daily data were recalculated on a monthly average.

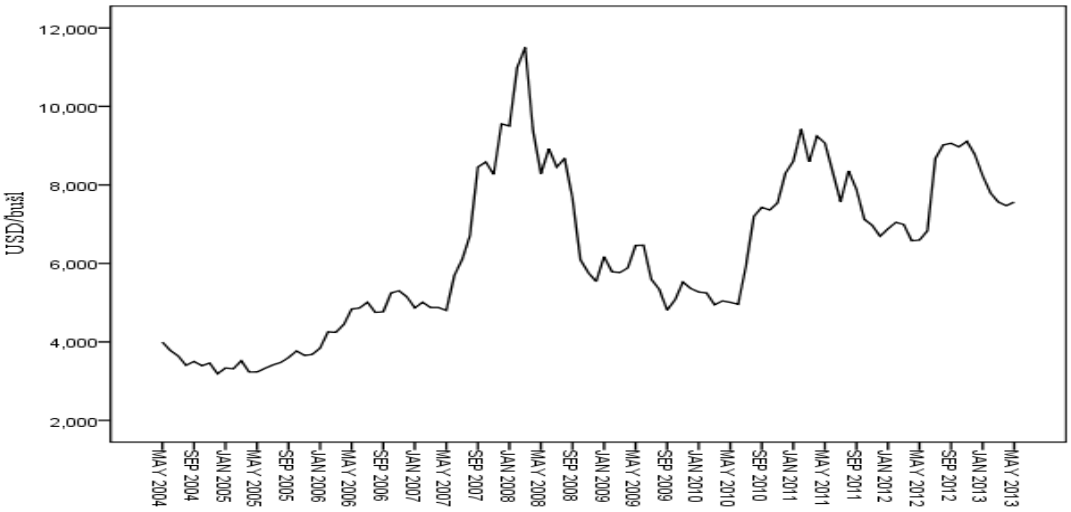


Figure 1 Wheat prices

[1] School of Business Administration in Karviná, Silesian University in Opava, Department of Mathematical Methods in Economics, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech Republic, stoklasova@opf.slu.cz

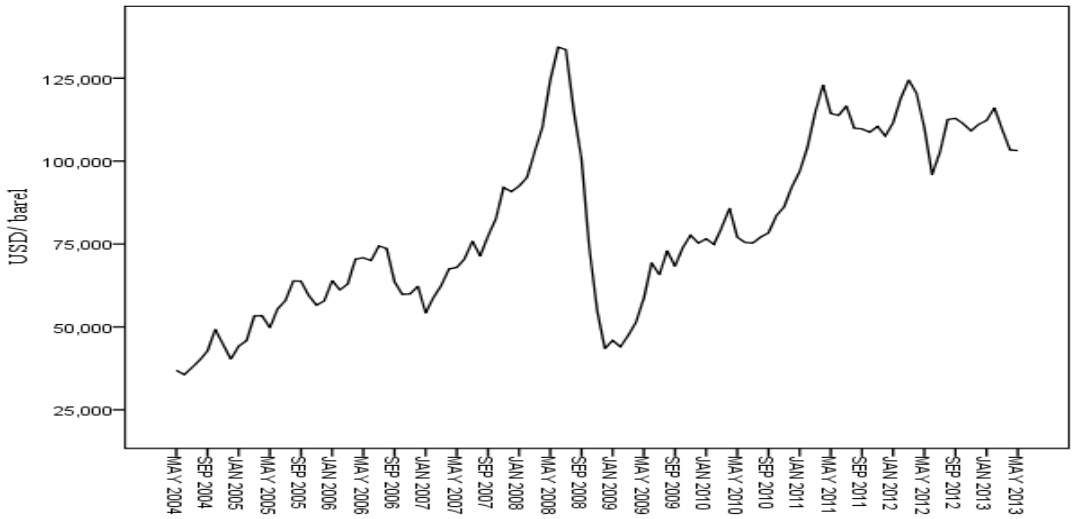


Figure 2 Oil prices

Comparing the minimum and maximum prices for wheat we can find that the price of wheat has increased over the period by 361%. The minimum price of wheat was in December 2004, and the maximum value was reached in March 2008, when the world was hit by the food crisis, which caused a domino effect on the commodity exchanges.

The price of oil increased during the period by 377%. The lowest price of oil was reached in June 2004 and the highest in June 2008, which was caused by supply shortages of oil from OPEC countries. The financial crisis caused decrease in the price of oil at the end of 2008.

Highest prices of observed commodities were recorded in the same year, which may mean that there is a correlation between the prices of the mentioned commodities.

	Wheat	Oil
N	109	109
Minimum	3.186	35.584
Maximum	11.504	134.353
Mean	6.22663	80.39665
Std. Deviation	2.034483	25.987648
Skewness	.278	.198
Kurtosis	-.872	-1.112

Table 1 Descriptive characteristics of wheat and oil prices

The value of the correlation coefficient 0.81 indicates that there exists high dependence between the prices of oil and wheat. The coefficient is statistically significant at the 0.01 level of significance. The high value of the correlation coefficient, however, still does not confirm the existence of a causal relationship.

		Wheat	Oil
Wheat	Pearson Correlation	1	.808**
	Sig. (2-tailed)		.000
Oil	Pearson Correlation	.808**	1
	Sig. (2-tailed)	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

Table 2 Correlation coefficient

2 Linear regression

Using linear regression we investigate the relationship of independent random variables *x* (oil prices) and the dependent variable *y* (wheat prices). Simple linear regression model is given by the equation:

$$y = a + bx + e, \tag{1}$$

where *a* is absolute term, *b* je regression coefficient a *e* is residual deviation.

We will check whether the time series are stationary, so that there can be applied regression analysis. In [1], [3] we can see that stationarity of the time series is tested using the extended Dickey-Fuller test (ADF test), which also detects whether a time series contains a unit root. For testing the presence of a unit root there are considered three different regression equation:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t, \tag{2}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t, \tag{3}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t. \tag{4}$$

If $\gamma = 0$, the regression coefficient is statistically insignificant and the time series contains a unit root. Stationarity of the time series is assessed by the t-statistic, which is compared with the critical value of Dickey-Fuller table. If the t-statistic is less than the critical value, the time series is stationary and there can be uses the linear regression model.

Data	Test Statistics	Critical Value (α =0,05)	Result
Wheat	– 2.01	– 2.86	Nonstationary TS
Oil	– 2.46	– 2.86	Stationary TS

Table 3 ADF – Original time series

One of the methods to obtain stationary time series is differentiating of time series (first difference). Again, there is applied an expanded Dickey-Fuller test.

Data	Test Statistics	Critical Value (α =0,05)	Result
Wheat	– 3.71	– 2.86	NonstationaryTS
Oil	– 6.58	– 2.86	Stationary TS

Table 4 ADF – Differentiated time series

Simple linear regression is performed for differentiated time series.

First, the case that the price of wheat is dependent on the price of oil is considered. The calculated regression model can be expressed as y =1,141 + 0,063x. The value of the coefficient of determination is 0.808; regression coefficient is statistically significant at the 0.01 level of significance.

Now consider that the oil price is dependent on the price of wheat. The regression model can be described by the equation y = 16,126 +10,322x. The value of the coefficient of determination is 0.808; regression coefficient is statistically significant at the 0.01 level of significance.

3 EC model

Error correction model is suitable for establishing long-term and short-term relationships. Using the model there were identified parameters that characterize the degree of deviation from the equilibrium relationship. The coefficients were calculated using the equation

$$\Delta Y_t = \beta_0 + \beta_1 \Delta X_t + \beta_2 res_{t-1} + \varepsilon_t. \quad (5)$$

First, again considering a model where the wheat is dependent on oil and the model equation is:

$$\Delta Y_t = 0{,}02132 + 0{,}03242 \Delta X_t - 0{,}07311 res_{t-1} + \varepsilon_t.$$

Given that the coefficient is not statistically significant, it cannot be said that there is a short-term or long term relationship.

In the latter case, when considering that the oil is dependent on wheat, the values of the coefficients are statistically significant at the 5% significance level. The equation of that model is as follows:

$$\Delta Y_t = 0{,}61241 + 2{,}71331 \Delta X_t - 0{,}13212 res_{t-1} + \varepsilon_t.$$

The equation shows that 13.2% of today's price increase will be adjusted in the next month and almost completely absorbed after 8 months. There is thus both a short-term relationship between the price of wheat and the price of oil, as well as a long-term relationship. Next information we can see in [8].

4 Granger causality

Correlation analysis determined the dependency rate of time series, but it did not specify which variable is the cause and which the effect. Using Granger causality there can be found short-term dependency and causality, we can see this in [6], [7]. It is believed that the price of wheat is prior to the price of oil, because it is a part of it. When testing Granger causality there is necessary that the investigated time series are stationary. Thus, differentiated time series are tested. First, there is determined the size of the delay using the VAR system, where it was found that the prices of commodities precedes by one month. Using vector autoregression there is determined whether the price of oil precedes the price of wheat.

	Probability							
Delay	1	2	3	4	5	6	7	8
d_oil→d_wheat	0.64	0.42	0.23	0.35	0.61	0.51	0.81	0.39
d_wheat→d_oil	0.01	0.18	0.31	0.75	0.68	0.21	0.39	0.29

Table 5 Testing Granger causality

The table shows that the change in the price of wheat precedes by one month the change in the price of oil, which contradicts the above assumption. “d_oil→d_wheat” means delay. There was identified the cause and its consequence: if the price of wheat increases (decreases) is one month then the price of oil increases (decreases) in the following month. It was therefore established a short-term relationship between the commodities.

5 Conclusion

The aim of the article was to determine the relationship between the price of oil and the price of wheat. The value of the correlation coefficient shows a strong dependence.

In the case of linear regression, where wheat was regarded as a commodity, which depends on the price of oil, it was found that the increase in oil prices of 1 USD per month, the price of wheat increases by 0.063 USD. Otherwise - if the price of wheat goes up by 1 USD per month then the price of oil increase by 10.322 USD. It can therefore be argued that there is a linear relationship between commodity prices.

When examining the long-term relationship with the linear regression residue there was found that the increase in the price of wheat in a given month is absorbed from 13.2% in the next month, and the price goes back to the equilibrium state after 8 months.

Granger causality rejected the argument that the increase in oil prices is further reflected in the price of wheat. Test showed that the change in the price of wheat precedes by one month the change in oil prices.

Acknowledgements

This paper was supported by the project "Innovation of Educational Programs at Silesian University, School of Business Administration in Karviná" n. CZ.1.07/2.2.00/28.0017.

References

[1] Artl, J.: *Moderní metody modelování ekonomických časových řad*. Praha, GRADA Publishing, 1999.
[2] Askari, H.: *The sweel and sour of oil*. The Asia Times. 31.May 2012. http://www.atimes.com/atimes/Middle_East/NE31Ak02.html.
[3] Cipra, T.: *Finanční ekonometrie*. 1.vyd. Praha: Ekopress, 2008.
[4] Devold, H.: *Oil and gas production handbook.* An introduction to oil and gas production. 2.3.ed. Oslo: ABB AS, 2010.
[5] Ghoshray, A.: *Price linkages in the international wheat market.* http://ecsocman.hse.ru/data/501/664/1219/090.pdf. 2003
[6] Hamilton, J.: *Time Series Analysis*. Princeton: Princeton Univ.Press, 1994.
[7] Mate, J.G.V. and Gilbreath, G.H.: *Statistics fo Business and Economics*. Third Edition. Homewood Illinois, 1987.
[8] ŠPERKA, R., SPIŠÁK, M. *Transaction Costs Influence on the Stability of Financial Market: Agent-based Simulation*. Journal of Business Economics and Management, Taylor & Francis, London, United Kingdom, 2013. Volume 14, Supplement 1, pp. S1-S12, DOI: 10.3846/16111699.2012.701227. Print ISSN 1611-1699, Online ISSN 2029-4433. Available from: <http://www.tandfonline.com/doi/abs/10.3846/16111699.2012.701227#.Ur80j9LuLy0>.

On the measurement of sustainable economic performance

Luboš Střelec[1], David Hampel[2], Ladislava Issever Grochová[3], Jitka Janová[4]

Abstract. As often discussed GDP is an inadequate metric to capture well-being, particularly, in its economic-social and environmental dimensions, aspects referred to as sustainability. In order to measure sustainable economic performance, it is necessary to reflect both economic and environmental aspects. Alternative indicators are often used for such a measure, taking GDP as a base and modifying it for its shortcomings. These alternative indicators take into account the environmental damage and the exploitation of natural resources. As these individual indicators are, however, constructed in various ways, this can lead to misunderstanding of the indicator content and hence to their misuse. Therefore, it is necessary to verify their performance and examine relationships among them which is the goal of this paper. Moreover, these indicators will be used to find potential differences in sustainable economic performance between EU old and new member states. For this purpose, a segmentation of the EU countries using cluster analysis according to their emphasis on economic development and environmental protection as well as the dynamic analysis of the segmentation is provided.

Keywords: Alternative economic indicators, cluster analysis, EU28, sustainable economic performance.

JEL Classification: C38, E01, Q01, Q32
AMS Classification: 62H30, 91C20, 91G70

1 Introduction

Today's fast-moving world is addressing long-run challenges as globalization and limited quantum of resources. These tasks can be fulfilled by many ways, but all of them will require information about individual subjects, typically countries. Researchers wish to select leading countries and determine a 'distance' among them; also a 'measure' of particular differences is a desirable characteristics (see e.g. [5] and [8]). Generally, an attention is paid to measurability of economic-environmental performance of the countries.

The endeavour to measure economic performance originated in the post war period of reconstruction. Many people lived in conditions of misery caused by destructive war and the key to prosperity was seen in greater production. As a result gross domestic product was regarded as the main indicator for measuring production and consequently its growth. This indicator has been implemented in economic practice by early thirstiest by S. Kuznets on direct request of American government. It main purpose was to measure the gross output of an American economy, particularly its production capacities in strategic industries.

This indicator has been frequently interpreted as a measure of the society´s well-being and society´s social progress measure. All this has continued despite the warning from the author of GDP on its unsuitability [7] for such applications. Experience shows that the broad applications and categorical interpretation of this indicator is not appropriate and do not give a comprehensive picture of the development of the society, neither on the social status of the population and the state of the environment. Simultaneously, many economists have pointed out that GNP and GDP can give a highly misleading impression of economic and human development [1] and [2].

With increasing requirements on capturing economic growth in all its complexity raised demand for a new concept referring to an overall condition. New terms as a 'well-being' and 'sustainable development' started to be used and of course to be measured.

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, 613 00, lubos.strelec@mendelu.cz.
[2] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, 613 00, qqhampel@mendelu.cz.
[3] Mendel University in Brno, Faculty of Business and Economics, Department of Economics, Zemědělská 1, Brno, 613 00, ladislava.grochova@mendelu.cz.
[4] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, 613 00, jitka.janova@mendelu.cz.

A number of ways of measuring national-level progress have been proposed, developed, and used to address sustainable development or less general areas. In general, these new measures can be categorized as:

1. indices that address the issues described above by making 'corrections' to existing GDP and SNA accounts (e.g. Human Development Index (HDI), Genuine Progress Indicator (GPI), Green GDP, Genuine savings, etc.);
2. indices that measure aspects of well-being directly (e.g. Ecological Footprint (EF), Environmental Performance Index (EPI), Subjective well-being, Gross National Happiness, etc.);
3. composite indices that combine the aforementioned approaches (e.g. Better Life Index (BLI), Living planet report, Happy planet index (HPI), etc.) and
4. indicator suites (e.g. National income satellite accounts, Calvert-Henderson quality of life indicators, Millennium development goals and indicators, etc.).

A question is whether GDP should be improved on, being replaced by these other approaches, or supplemented. Advocates for supplementing GDP with these other measures point out that GDP, while being a poor measure of welfare, nonetheless "serves crucial and helpful roles in macroeconomic policy" and is "unique in that it combines simplicity, linearity, and universality as well as carries the objectivity of the observable market price as its guiding principle" [4].

As stated in [9], measuring better lives has become even more important today, as many of our economies and societies have been stricken by the global financial or economic crisis. Understanding how the lives of people have been affected and designing the best strategies to help them seems to be a crucial task. It is thus important to have as accurate as possible information on how both people's economic and noneconomic well-being have evolved during the crisis. Well-being metrics can provide a new and wider perspective to policymakers in the areas that matter to people. Such expanded set of indicators can also open new horizons in traditional policy areas by providing a new type of information, such as how people behave and feel about their lives [9]. For correct usage of well-being indicators it is necessary to explore them in detail.

The aim of this contribution is then to present and evaluate a set of tools that allow to measure and to evaluate not only the results of economic growth, but also its complex social and environmental impacts. The rest of the paper is organized as follows: section 2 includes description of used indices and methodology, section 3 dissects results and the last section concludes.

2 Empirical study

For our analysis we use the following indicators reflecting the economic, environmental and sustainable development performance[5]: Gross Domestic Product in PPS *per capita* (GDP), Human Development Index (HDI), Better Life Index (HPI), Well-being indicator (WB), Ecological Footprint (EF) and Environmental Performance Index (EPI). Mentioned indicators are described in detail in [6]. For all calculations and graphical outputs R software is used [10].

Tab. 1 presents economic, environmental and sustainable indicators of EU countries for 2012. As can be seen, the core countries have the highest values of the indicators, as the well-being indicators are accompanied with high ecological burden. In contrast, the lowest values of indicators analyzed have Central and Eastern Europe countries (CEEC) mainly due to Bulgaria and Romania.

2.1 Relationships between indicators

Tab. 2 and 3 report the Pearson and Spearman correlation coefficients of the mentioned indicators and GDP for the years 2010 and 2012. Relationship between the abovementioned indicators and GDP in PPS *per capita* for the year 2012 are presented in Fig. 1 and 2. These tables and figures imply a strong relationship between the indicators and GDP in PPS *per capita*. The weakest relationship is evident between GDP and EPI, and GDP and HPI. In general, it can be stated that the correlation coefficients are mostly statistically significant. The highest values of the Pearson correlation coefficients between GDP and HDI, GDP and BLI, and between GDP and WB support also the theoretically formulated expectation as the indicators are derived from GDP. On the other hand, no Pearson correlation for the year 2010 is observed between HDI and HPI, BLI and EF, HPI and EF, HPI and EPI, WB and EPI, and between EF and EPI. Unlikely in 2010, in 2012 no Pearson correlation is observed only between HPI and EF, HPI and EPI, and between EF and EPI. These findings are even reinforced by Spearman correlation coefficients.

[5] Sources and indices definitions can be found here: Eurostat; United Nations Development Programme; Centre for Well-being at nef (the new economics foundation); Global Footprint Network, 2012, National Footprint Accounts, 2012 Edition; Yale University (Yale Center for Environmental Law and Policy).

	GDP	HDI	BLI	HPI	WB	EF	EPI
Austria	33100	0.895	6.65	47.09	7.35	5.29	68.92
Belgium	30700	0.897	6.70	37.09	6.85	7.11	63.02
Bulgaria	12100	0.782	.	34.15	4.22	3.56	56.28
Croatia	15700	0.805	.	40.62	5.60	4.19	64.16
Cyprus	23400	0.848	.	45.51	6.39	4.44	57.15
Czech Republic	20700	0.873	5.31	39.35	6.15	5.27	64.79
Denmark	32100	0.901	7.07	36.61	7.77	8.25	63.61
Estonia	18200	0.846	4.86	34.95	5.14	4.73	56.09
Finland	29400	0.892	7.05	42.69	7.39	6.21	64.44
France	27700	0.893	6.26	46.52	6.80	4.91	69.00
Germany	31500	0.920	6.67	47.20	6.72	4.57	66.91
Greece	19500	0.860	4.88	40.53	5.84	4.92	60.04
Hungary	17000	0.831	4.67	37.40	4.73	3.59	57.12
Ireland	32900	0.916	7.15	42.40	7.26	6.22	58.69
Italy	25600	0.881	5.54	46.35	6.35	4.52	68.90
Latvia	16400	0.814	.	34.87	4.67	3.95	70.37
Lithuania	18300	0.818	.	34.55	5.07	4.38	65.50
Luxembourg	67100	0.875	6.55	28.99	7.10	10.72	69.20
Malta	22000	0.847	.	43.10	5.77	4.26	48.51
Netherlands	32600	0.921	6.98	43.09	7.50	6.34	65.65
Poland	17100	0.821	4.95	42.58	5.78	3.94	63.47
Portugal	19400	0.816	5.25	38.68	4.87	4.12	57.64
Romania	13500	0.786	.	42.18	4.91	2.84	48.34
Slovakia	19400	0.840	5.11	40.13	6.05	4.66	66.62
Slovenia	21400	0.892	6.02	40.17	6.08	5.21	62.25
Spain	24400	0.885	6.27	44.06	6.19	4.74	60.31
Sweden	32200	0.916	7.47	46.17	7.50	5.71	68.82
United Kingdom	26800	0.875	6.89	47.93	7.03	4.71	68.82
EU	25007	0.862	6.11	40.89	6.18	5.12	62.67
V4	18550	0.841	5.01	39.87	5.68	4.37	63.00
CEEC	17255	0.828	5.15	38.27	5.31	4.21	61.36
PERIPHERY	24360	0.872	5.82	42.40	6.10	4.90	61.12
CORE	34320	0.899	6.83	42.34	7.20	6.38	66.84

Table 1 Economic, environmental and sustainable indicators of the EU countries groups for 2012[6]

[6] Countries: **V4:** Czech Republic, Hungary, Poland, Slovakia; **CEEC:** Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia; **Periphery:** Greece, Ireland, Italy, Portugal, Spain. **Core:** Austria, Belgium, Denmark, Finland, France, Germany, Luxembourg, Netherlands, Sweden, United Kingdom.

	GDP	HDI	BLI	HPI	WB	EF	EPI
GDP		0.95*	0.85*	0.42*	0.92*	0.64*	0.35
HDI	0.94*		0.79*	0.42*	0.86*	0.63*	0.35
BLI	0.91*	0.85*		0.60*	0.89*	0.23	0.56*
HPI	0.40*	0.32	0.65*		0.49*	-0.21	0.22
WB	0.92*	0.87*	0.92*	0.50*		0.61*	0.25
EF	0.64*	0.66*	0.29	-0.35	0.63*		0.07
EPI	0.39*	0.43*	0.60*	0.13	0.27	0.15	

Table 2 Pearson (below the diagonal) and Spearman (above the diagonal) correlation, analyzed year 2010 (HPI, WB and EF: 2009); * $p<0.05$
Note: correlation coefficients are computed based on the dataset without Luxembourg due as it is an outlier in GDP.

	GDP	HDI	BLI	HPI	WB	EF	EPI
GDP		0.95*	0.89*	0.58*	0.93*	0.82*	0.41*
HDI	0.93*		0.84*	0.51*	0.90*	0.86*	0.38*
BLI	0.94*	0.86*		0.37	0.93*	0.72*	0.39
HPI	0.54*	0.52*	0.46*		0.57*	0.19	0.38
WB	0.93*	0.89*	0.91*	0.60*		0.85*	0.44*
EF	0.78*	0.75*	0.68*	0.02	0.80*		0.30
EPI	0.44*	0.45*	0.46*	0.26	0.46*	0.33	

Table 3 Pearson (below the diagonal) and Spearman (above the diagonal) correlation, analyzed year 2012; * $p<0.05$
Note: correlation coefficients are computed based on the dataset without Luxembourg due as it is an outlier in GDP.

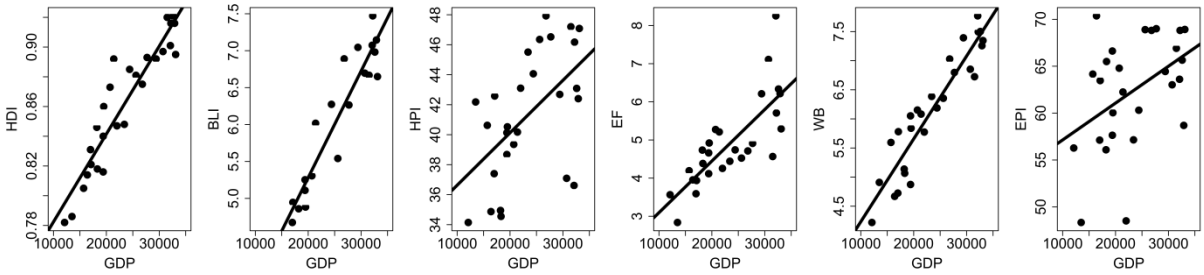


Figure 1 Relationship between indicators and GDP in PPS *per capita*, analyzed year 2012
Note: Constructed scatter plots are based on the dataset without Luxembourg due to outlier in GDP.

2.2 Cluster analysis

In order to explore dis/similarity of the CEE countries and its changes over time, cluster analysis is applied. For detailed discussion on cluster analysis and its use see, for example [3]. In particular, we employ Ward method and Euclidean distance. Fig. 3 presents dendrograms which are based on the selected environmental and sustainable indicators for analyzed years 2010 and 2012 that show the dissimilarity of the EU members' attitude to the environment. A special attention in the comparative analysis of the environmental and sustainable indicators is given to their ability to classify the individual countries/economies into generally accepted homogenous groupings. It is evident that there is a clear segmentation of the EU members into two main clusters. The first cluster, in general, consists of the core states plus better performing periphery and CEE countries. In 2010, this cluster includes beside the core countries also Ireland, Spain, Greece, Italy, Cyprus, the Czech Republic and Slovenia. While in 2012, the described cluster includes the core countries and Ireland and Italy. The other mentioned countries have moved to the second cluster, which contains mostly the CEE countries.

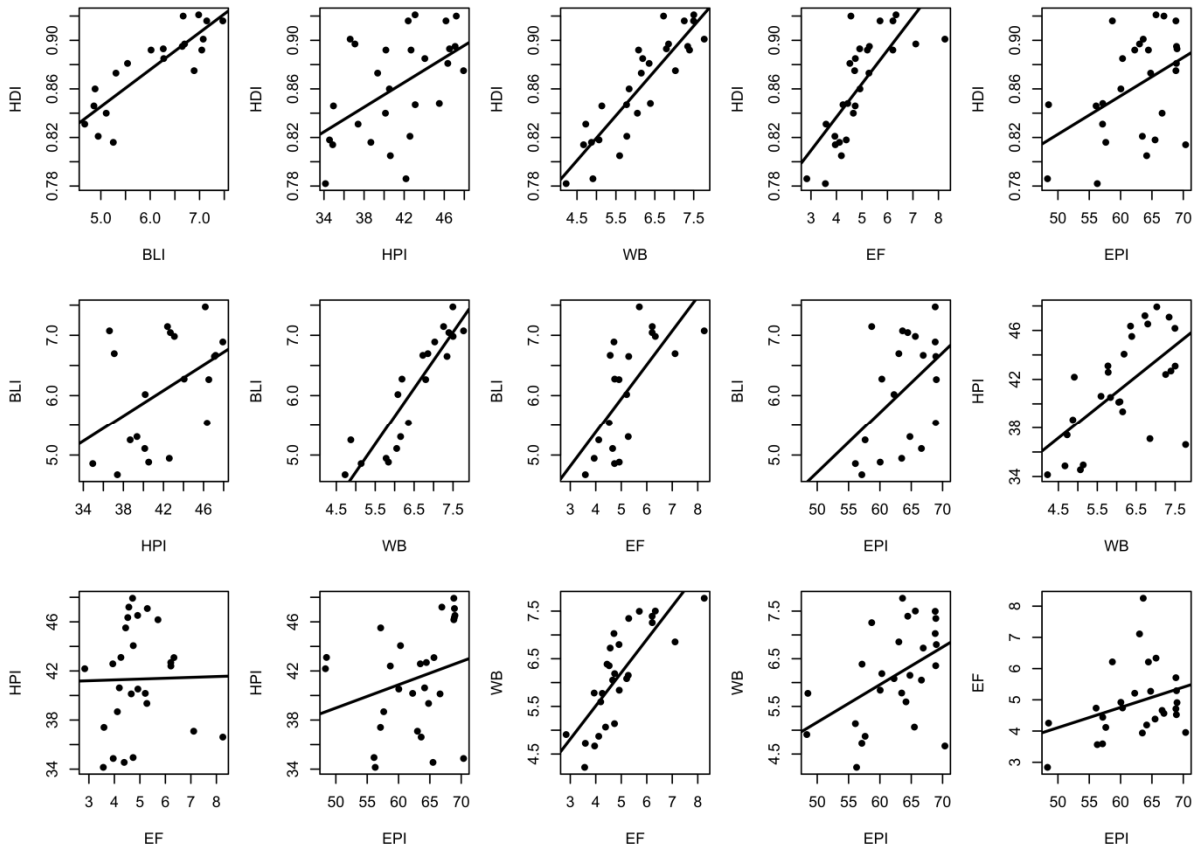


Figure 2 Relationship between indicators, analyzed year 2012

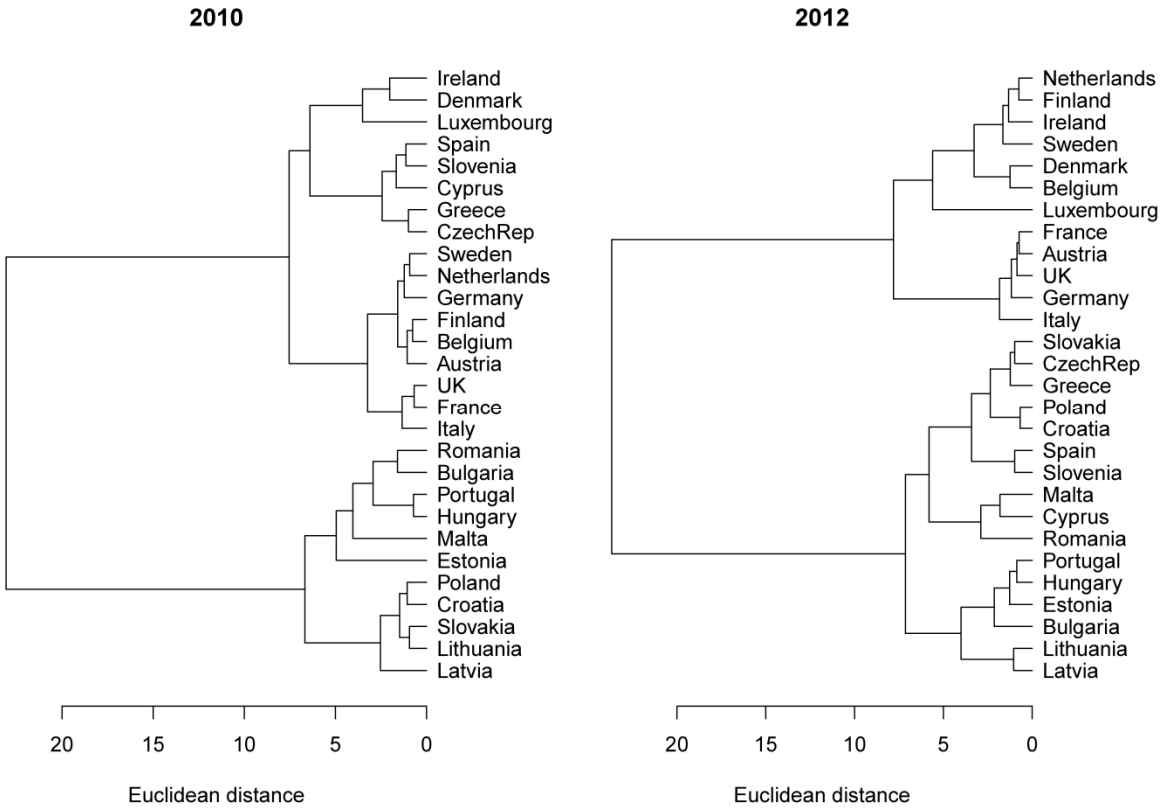


Figure 3 Dissimilarity of the EU members in environmental and sustainable indicators, analyzed year 2010 (HPI, WB, EF: 2009)

Note: Dendrograms are based on following indicators: HDI, BLI, HPI, WB, EF, and EPI; for its construction are used Ward method and Euclidean distance.

3 Conclusions

The contribution introduced the development of the content of various measures of sustainable economic performance. The alternative indicators of sustainable economic performance, which take into account the environmental damage and the exploitation of natural resources, were presented. From our analysis can be seen the strong relationship between analyzed indicators and GDP in PPS *per capita*, especially between GDP and HDI, GDP and BLI, and between GDP and WB. We also used these indicators to find potential differences in sustainable economic performance between EU old and new member states. In fact, the core countries when compared to the CEE countries, generally demonstrated higher values of all studied indicators. Based on cluster analysis the EU countries can be divided into two main clusters. The first cluster consists primarily of the core countries plus better performing periphery and CEE countries. From the dynamics of the segmentation can be observed a shift of most of these periphery and CEE countries to the second cluster and formation of a homogeneous cluster consisting primarily of old EU countries with the exception of countries of Mediterranean enlargements, i.e. Greece, Spain and Portugal.

Acknowledgements

The contribution is partially supported by the grant No. 13-25897S of the Czech Science Foundation. We also thank to Prof. Kabát, whose insightful comments helped us to improve the paper considerably.

References

[1] Bell, S., Morse, S.: *Measuring sustainability: Learning from doing*. Earthscan, 2003.
[2] Bell, S., Morse, S.: *Sustainability Indicators: Measuring the Immeasurable?* Earthscan, 2008.
[3] Everitt, B. S., Landau, S. and Leese, M.: *Cluster Analysis*, Oxford University Press, Inc., New York, 2001.
[4] Goossens, Y., Mäkipää, A.: Alternative Progress Indicators to Gross Domestic Product (GDP) as a Means Towards Sustainable Development. *Brussels: European Parliament, Policy Department A: Economic and Scientific Policy* (2007).
[5] Grochová, L., Rozmahel, P.: Business cycle correlation of the CEEC and the Euro area: some methodological controversy. In: *Proceedings of 30th International Conference on Mathematical Methods in Economics* (Ramík, J., Stavárek, D.), Karviná, 2012, 237–242.
[6] Kabát, L., Hampel, D., Issever Grochová, L., Janová, J., Střelec, L.: Alternative approaches for assessing the European countries economic and social results. Submitted to *Procedia Economics and Finance* (2014).
[7] Kuznets, S.: *National Income 1929–1932. A report to the U.S. Senate*, 73rd Congress, 2nd Session. Washington, DC. US Government Printing Office, 1934.
[8] Lawn, P. A.: A Theoretical Foundation to Support the Index of Sustainable Economic Welfare (ISEW), Genuine Progress Indicator (GPI), and Other Related Indexes, *Ecological Economics* **44** (2003), 105–118.
[9] OECD 2013. *How's life? 2013: Measuring well-being*. OECD Publishing, 2013.
[10] R Core Team: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014, http://www.R-project.org/.

Analysis of Portfolio Options in the Field of Financial Logistics

Petr Suchánek[1], Franciszek Marecki[2], Monika Marecka[3], Robert Bucki[4]

Abstract. Financial logistic analysis is based on the flows of risky capitals which consist of complexes of financial operations. Management of risky capital or, in general, risk management in the complex is a real and current topic of contemporary logistics. The risk management of risky capitals consists of solving dual optimization problems associated with profit maximization of a given risk and risk minimization at a given profit. Optimization is based on the analytical methods which is emphasized by the analysis of combinatory Call and Put options. Analysis of options is based on mathematical models of nonlinear time-varying rates and profits, hence requiring in practice the use of computer programs. The main essence of Call and Put options is the identification of static characteristics as well as the price and borderline course of these options. These characteristics can be determined recursively, updating the database with concluded contracts for options. The use of static characteristics allows us to derive the general functions of profits from options. The analysis is based on the principle of equivalence of capitals.

Keywords: financial logistics, risky capital, Call option, Put option, optimization, risk management.

JEL Classification: G11, C02, C61
AMS Classification: 90A09, 65C20

1 Introduction

Financial logistics is a base of classical logistics consisting of production, transport, storing, distribution, information flows, etc. Financial logistics analysis streams of risky capitals and consists of complex of financial operations. Risk is a characteristic of capital investments which consist of the use of currently available resources for the probable future profits [2], [5], [11], [12], [15]. In general, the measurement of risk is the probability of achieving those future profits. In practice, investments offering higher returns are riskier (because higher returns are less likely). The risk management consists of solving of dual optimization problems associated with maximization of profit for a given risk and minimization of risk at a given profit. Risk management requires the use of computing systems for recording of market data and support of financial decision-making [3]. In the general case this constitutes a complex problem that can be solved using a variety of methods presented in e.g. [7], [9], [10], [16]. In this article, the method called Mark-to-Future (MTF) [4], [6], [13] is taken into account. The concept of this method is to highlight standard financial instruments, identify investment scenarios, adopt the investments standards and conduct repetitive calculations of risk and return parameters for all variants of any financial instrument. On the basis of such a database one can create investment packages by aggregating the standard financial instruments from the database. For these packages there are calculated risk and return parameters that allow to manage the risk of the investment process. This modern, complex approach is costly; hence it is mostly used for high-value investments. The risk of the investments may be constrained by different types of insurance. In the case of the capital markets the so-called derivatives such as options are used as insurance [8]. An option is a share used for insuring capital market instruments such as currencies, stocks, bonds, etc. A characteristic feature of these instruments is their variable rate or the market price. The purpose of an option is to insure an investor against excessive increase (or decrease) of the instrument price. The article emphasizes the analysis of options with the use of the insuring currency example.

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Informatics, Univerzitní náměstí 1934/3, 73340 Karviná, Czech Republic, suchanek@opf.slu.cz.
[2] Wyższa Szkoła Biznesu w Dąbrowie Górniczej, ul. Cieplaka 1C, 41-300 Dąbrowa Górnicza, Poland, fmarecki@gmail.com.
[3] Institute of Management and Information Technology in Bielsko-Biała, ul. Legionów 81, 43-300 Bielsko-Biała, Poland, mmarecka@gmail.com.
[4] Institute of Management and Information Technology in Bielsko-Biała, ul. Legionów 81, 43-300 Bielsko-Biała, Poland, rbucki@wsi.net.pl.

2 Risk analysis

Analysis of options is based on mathematical models of nonlinear time-varying rates and profits, hence requiring in practice the use of computer programs [14], [1]. The issue of investment risk analysis is presented in two examples that justify the need and the possibility of insurance through options. Let us consider the first case of an investor who, in time $t = 0$, has signed a contract for the purchase of a foreign machine where the price V is expressed by means of the currency (e.g. €). In addition, the contract specifies the long lead time for payment, (e.g. T). The investor may adopt the following two strategies for the implementation of this contract:

1) Make a money deposit V_0 for a period [0, T] so that (1):

$$V = V_0 \cdot \prod_{t=1}^{T}(1+r_t) \quad \left(V_0 = \frac{Z_0}{q_0}\right) \tag{1}$$

where : q_0 – the currency exchange rate at the time t=0; r_t – the interest rate of the money deposit at the time t; Z_0 – the amount (in PLN).

2) Use the amount Z_0 for the economic activity in the period [0, T] which results in the return of the investment Z_T.

If the condition (2):

$$V > \frac{Z_T}{q_T} \tag{2}$$

holds, the first strategy is more profitable as the amount Z_T (in PLN) is insufficient to cover the purchase price V of the machine. The problem consists in the increase of the exchange rate q_T. Despite this fact, companies typically spend their owned capitals Z_0 on an economic activity (which is their goal) rather than on bank deposits. Besides, in case when the condition (2) could not be met, the second strategy (an economic activity) would be more profitable. For this reason, the company should be insured against the case when the exchange rate rises, and as a result, the condition (2) is satisfied. Note that the rate may increase but it cannot exceed the limit g_c fixed by the formula (3):

$$g_C = \frac{Z_T}{V} \tag{3}$$

An insurance policy against the increase of the price (when purchasing a foreign currency) is a CALL type option. Using such a policy, a company can purchase the currency at an exchange rate g_c. Let us consider an investor who in the time $t = 0$ received an amount V_0 in a foreign currency (e.g. €). In the end-of-year balance sheet (at the time T) this capital should be expressed in the local currency (e.g. in PLN). The investor may consider one of the two following investment strategies:

1) Make a foreign currency money deposit V_0 for the period [0, T] so that (4):

$$V_T = V_0 \cdot \prod_{t=1}^{T}(1+r_t) \tag{4}$$

which at the end of the year has a value of $V_T \cdot q_T$ expressed in PLN
where: q_T – exchange rate at time T.

2) Use the amount $Z_0 = V_0 \cdot q_0$ for an economic activity in the period [0, T], resulting in the return on the investment amount Z_T (e.g. at the end of the year).

If the condition $V_T \cdot q_T < Z_T$ holds, the second strategy is more profitable.

Companies typically compare their foreign capital V_0 for their economic activity (which is their primary purpose) rather than for bank deposits. If, however, the company establishes a foreign currency bank deposit, it should be insured against the case when the condition $V_T \cdot q_T < Z_T$ is satisfied. The problem of foreign bank deposits is caused by possible decreases in rate q_T. Note that the rate could rise but, despite this, the second strategy will be more profitable. The limiting value of the course g_P is determined from the formula (5):

$$g_P = \frac{Z_T}{V_T} \tag{5}$$

Insurance policy for the decline in the course (when selling the currency) is a PUT type option. On the basis of such a policy the company may sell the currency at the borderline exchange rate g_P.

3 The CALL option

Let us consider the currency exchange rate fluctuation in the period [0, T]. As it can be seen in Figure 1, the exchange rate has a tendency to rise from q_0 to q_T.

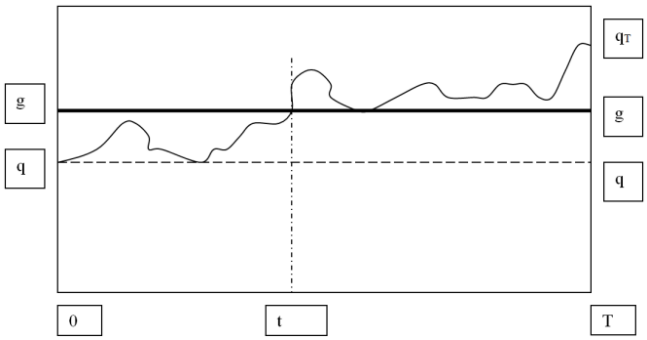


Figure 1 The process of the increase of the exchange rate value

At the time t, the exchange rate crosses the boundary value g. For such a case, an investor should purchase from a stock broker an insurance in the form of a CALL option. The option price C is an analogue of insurance premiums. The broker is obliged to sell the investor the currency at an exchange rate agreed upon in the option in times from t to T. The investor does not claim the right to purchase the currency if its exchange rate q_i does not exceed the boundary value g (for the currency is then less expensive at a market than at the stock broker's).

3.1 The price analysis of the CALL option

The problem of setting the price of options (CALL or PUT) is called the valuation of options. From a theoretical point of view it is a difficult problem of the analysis of a stochastic process which is the process of changing valuations. In practice, simplified statistical methods can be employed. In insurance policies, the principle of capital equivalence is often used, according to which "the premium C equals the probable discounted payment W" (6):

$$C = \frac{P_t \cdot W_t}{1 + s_t} \tag{6}$$

where P_t – the probability of a payment in the time t; W_t – the payment amount (difference : $q_t - g$) in the time t; s_t – the discount rate for the time period [0, t].

It follows from the equation (6) that the price C of an option depends on the boundary exchange rate g, the probability of exceeding the boundary exchange rate g and the state of the financial market expressed through an equivalent rate s_t. The value of the payment depends on the difference: $q_t - g$, yet greater differences are less likely. The probability of a payment depends on the duration of the insurance period – the longer the period, the more likely it is that the boundary exchange rate is exceeded. In addition, the breach of the boundary exchange rate depends on the currency exchange rate fluctuations which do have important implications. The payments are discounted as thus, their value at the time t=0 depends on the current market state (interest rates, inflation, etc.). In general, many forecasted and non-forecasted parameters affect the valuation of options and thus, the complex option valuation mathematical methods are often based on assumptions that are not met. For this reason, for the valuation of options a method of the identification of a static characteristic can be introduced: price – boundary exchange rate. Let us assume that a stock broker, after negotiations with a client, has issued N CALL options. The prices: $C_1,...,C_n,...,C_N$ and boundary prices: $g_1,...,g_n,...,g_N$ of the issued options are given. Employing the method of least-squares it is possible to calculate a static characteristic of options, from the equation $C = a \cdot g + b$ where a, b are constants calculated using the least-squares method from the data (C_n, g_n) for n = 1,...,N. In the general case of CALL options, the coefficient a is negative which means that the price of options decreases with the increase of the boundary interest rate $g > q_0$.

3.2 The profit of the CALL option

Insurance policies (for an investor - contributions) are costs and payments are incomes. If an insured event does not occur the investor suffers losses (incomes are negative contributions). In the opposite case - the income is the difference between the payment and the insurance premium. Investor's gain is the broker's loss - and vice versa.

Let us assume that the investor has purchased a CALL option by declaring to the broker the boundary rate g. The broker, using equation $C = a \cdot g + b$ has determined the price C of an option. Given that, the investor and the broker can calculate the characteristics of their profits. The investor calculates its profit Z from the conditions $Z = -C$ (for the exchange rate $q < g$) and $Z = -C + (q - g)$ (for the exchange rate $q > g$). In relation to the above conditions a function of the profit for the investor is obtained in the form (7):

$$Z = -C + \max(0, q - g) \tag{7}$$

As it can be seen in Figure 2 the profit for the investor is not bounded, albeit not likely for rates q>>g. Plugging equation $C = a \cdot g + b$ into the equation (7) the general profit function (8) is obtained:

$$Z = -a \cdot g - b + \max(0, q - g) \tag{8}$$

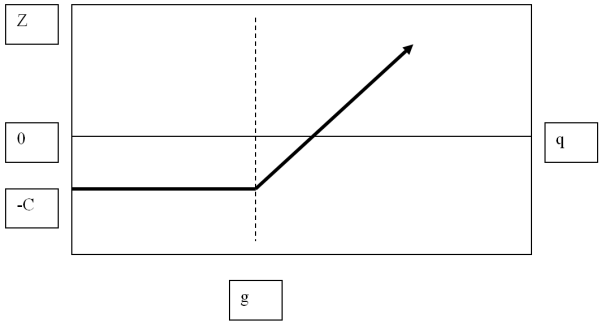


Figure 2 The characteristic of the profit from the CALL option for the investor

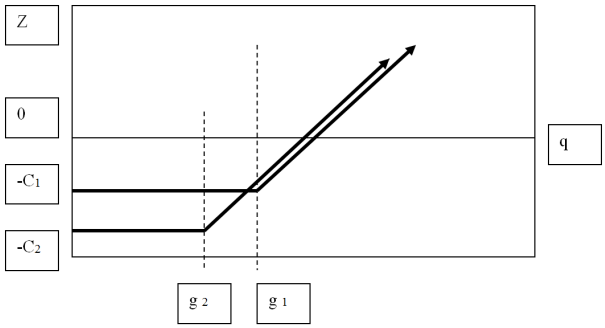


Figure 3 The general profit function for the investor from the CALL option

The general profit function for the investor from the CALL option shows a regularity consisting in the price of options being greater for smaller boundary rates. Moreover, as it can be seen, the difference of the option costs is big whereas the difference of the option returns is small (for a fixed rate). The characteristic of the profit for the broker is a mirror reflection (towards the q axis) of the characteristic for the investor (Figure 3).

4 The PUT Option

Let us consider the process of currency exchange rate changes during the interval [0, T]. As shown in Figure 4 the rate declines from q_0 to q_T.

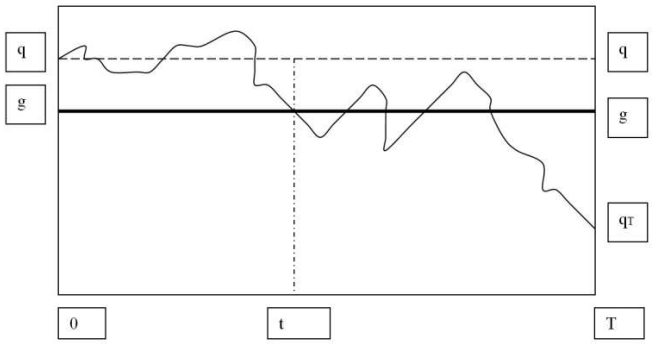


Figure 4 The decline process of the rate

In the time t the rate exceeds the limit g. In such a case an investor should purchase an insurance at the broker in the form of a PUT option. The price of the option C is an analogue of an insurance premium. The broker is obliged to buy the currency from an investor at the rate g fixed in the option in the period from t to T. The investor does not exercise his right to sell a currency at a broker, if its rate q_i is not lower than the limit value g (because the currency on the market is more expensive than while buying at a broker's office).

4.1 The price analysis of the PUT Option

The problem of setting the price for the PUT option is called the valuation of option. From a theoretical point of view, this is a difficult problem of analysis of the stochastic process - as for the CALL option. In practice, simplified statistical methods can be used such as the method of the least-squares. The principle of equivalence of capital (6) shows that the price C of the PUT option depends on boundary rate g (because the payment $W_t = g\text{-}q_t$), the likelihood of crossing the borderline exchange rate g and the state of financial markets as expressed by an equivalent rate of s_t. The value of the payment depends on the difference $g - q_t$, however, greater differences are less likely. The probability of payment depends on the duration of the insurance period. In addition, exceeding of the borderline rate depends on the currency fluctuation. The payments W_t are discounted so their value at the time $t = 0$ depends on the state of the market (discount rates, inflation, etc.). In general, many forecasted and non-forecasted parameters affect the valuation of the PUT option and thus, the complex option valuation mathematical methods are often based on assumptions that are not met. For that reason, for the valuation of a PUT option a method of the identification of a static characteristic can be employed: price – boundary exchange rate.

Let us assume that a broker, through negotiations with a client, issues N PUT options. The prices of the issued options are $C_1,...,C_n,...,C_N$ and the boundary rates are: $g_1,...,g_n,...,g_N$. Based on these data, using the method of the least-squares, the static characteristics of PUT option is set in the form of the equation $C = a \cdot g + b$. In the general case of the PUT option the "a" factor in the equation $C = a \cdot g + b$ is positive which means that the option price C increases with the growth of the borderline rate g while it is assumed that $g < q_0$.

4.2 The profit of the PUT option

In the general case of an insurance policy - for the investor - premiums are costs and payments are incomes. If an insured event does not occur the investor does not receive the payment and bears the loss (incomes are negative contributions). In the opposite case, when an insured event occurs, the investor's income is the difference between the payment and the premiums. In case of the insurance in capital markets, the profit of the investor is the loss for the broker and vice versa.

Let us suppose that an investor bought a PUT option and declared to the broker a border rate g. The broker, using the equation $C = a \cdot g + b$ **,** gives a price C of this option. On this basis, the investor (and the broker) may designate the characteristics of their gains. The investor designates his profit Z from the sets of conditions $Z = -C$ (for the exchange rate $q>g$) and $Z = -C + (g - q)$ (for the exchange rate $q<g$). Linking the above conditions the function of the investor gain (9) is obtained:

$$Z = -C + \max(0, g - q) \tag{9}$$

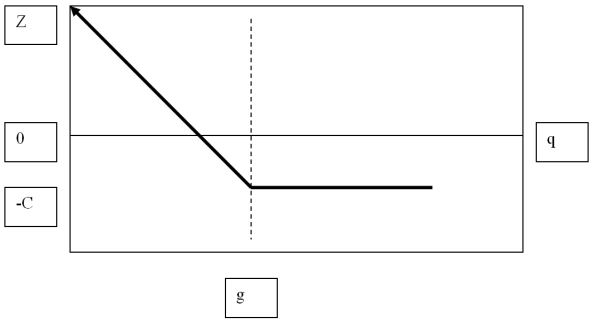


Figure 5 The characteristic of the profit of the investor from a PUT option

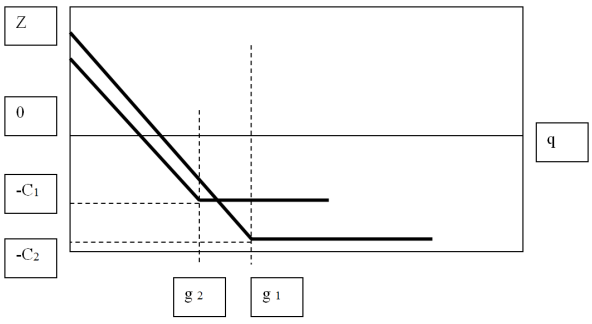


Figure 6 The overall function of the investor's profits from the PUT options

As shown in Figure 5 the maximum profit the investor is expected to get equals $g - C$, albeit this is not very likely for exchange rates $q << g$. The characteristics of the broker's profit from the PUT options is a mirror reflection (towards the q axis) of the one which is shown in Figure 5. Substituting $C = a \cdot g + b$ to (9) the general function of the investor's profits is obtained from the PUT options (10):

$$Z = -a \cdot g - b + \max(0, q - g) \tag{10}$$

The course of the profits for various borderline exchange rates is shown in Figure 6. The overall function of investor's profits from the PUT options shows the regularity that the option price is greater for larger borderline exchange rates. Furthermore, it can be noted that the difference between the options prices is big but the differ-

ence in profits is small (for a particular exchange rate). Due to the limited scope of the article, we can not present numerical example, but note that this theory has been verified and numerical example can be introduced, for example, in the extended version of the article in future.

5 Conclusions

The analysis of CALL and PUT options is based on the identification of static characteristics (price - borderline course) of these options. These characteristics can be determined recursively, updating the database with concluded contracts for options. The use of static characteristics (price - borderline course) allows us to derive the general functions of profits from options. Characteristics of the profit from options which come from the general functions of the profit from the CALL and PUT options have a different course than the graphs presented in the cited references in the bibliography. General functions of profits from CALL and PUT options can be easily used for the construction and analysis of option's portfolio of any composition. It enables us to manage the risk according to the MTF concept. According to the concept of this method we distinguish standard financial instruments - CALL and PUT options. For these options it is possible to specify scenarios for different periods and with different borderline courses. Furthermore, based on repetitively fixed static characteristics (price - borderline course) risk and profit parameters can be determined. This kind of option database enables the risk management of option portfolios for courses of any predictions.

Acknowledgements

The paper is one of the outcomes of the project "Innovation of Study Programs at Silesian University in Opava, School of Business Administration in Karviná" Nr. CZ.1.07/2.2.00/28.0017.

References

[1] Adam, L.: Stochastic programming software: a comparison for investment problem. In: *Proceedings of the 29th International Conference on Mathematical Methods in Economics*. Janska Dolina, 2011, 11-16.
[2] Amenc, N., and Martellini, L.: Risk Allocation: A New Investment Paradigm? *Journal of Portfolio Management*, **40**(2) (2014), 1–4.
[3] Borovicka, A.: The investment decision making under uncertainty. In: *Proceedings of 30th International Conference Mathematical Methods in Economics*. Karviná, 2012, 37-42.
[4] Brigham E. F., and Gapenski L. L.: *Zarządzanie finansami*. Polskie Wydawnictwo Ekonomiczne (PWE), Warszawa, 2000, 524-536.
[5] Coleman, T. S.: *A Practical Guide to Risk Management*. CFA Institute, Charlottesville, 2011.
[6] Dembo, R. S., Aziz, A. R., Rosen, D., and Zerbs M.: *Mark to Future – a Framework for Measuring Risk and Reward*. Algorithmics Publications, Toronto, 2000.
[7] Górecki, J., and Holeňa, M.: An Alternative Approach to the Structure Determination of Hierarchical Archimedean Copulas. In: *Proceedings of 31st International Conference Mathematical Methods in Economics*. Jihlava, 2013, 201–206.
[8] Hammoudeh, S., and McAleer, M.: Risk management and financial derivatives: An overview. *North American Journal of Economics and Finance*, **25** (2013), 109-115.
[9] Hu, Y. C.: Incorporating a non-additive decision making method into multi-layer neural networks and its application to financial distress analysis. *Knowledge-Based Systems*, **21**(5) (2008), 383-390.
[10] Kimmel, P. D., Weygandt, J. J., and Kieso, D. E.: *Financial Accounting: Tools for Business Decision Making*. Wiley, New York, 2011.
[11] Marecka, E., and Marecki, F.: *Modele i algorytmy analizy ryzyka finansowego ubezpieczeń życiowych*. WSB Dąbrowa Górnicza, 2011.
[12] Paye, B. S.: Elements of Financial Risk Management. *International Review of Economics & Finance*, **25** (2013), 451-452.
[13] Saunders, A., and Allen, L.: *Credit Risk Management In and Out of the Financial Crisis: New Approaches to Value at Risk and Other Paradigms*. Wiley, New York, 2010.
[14] Sperka, R., and Spisak, M.: Tobin tax introduction and risk analysis in the Java simulation. *Proceedings of 30th International Conference Mathematical Methods in Economics*. Karviná, 2012, 885-890.
[15] Yang, J.: Management of Individual Financial Risk. In: *Proceedings of the 5th International Conference on Financial Risk and Corporate Finance Management*. Dalian, 2013, 344-347.
[16] Zmeskal, Z.: Application of decomposition multi-attribute methods AHP and ANP in financial decision-making. In: *Proceedings of 6th International Scientific Conference on Managing and Modelling of Financial Risks*. Ostrava, 2012, 689-699.

Small open economy in period of recession: Nonlinear DSGE approach

Karolína Súkupová[1], Osvald Vašíček[2]

Abstract. In this paper, we study a model of a small open economy and its behavior with respect to the period of the global recession and the subsequent recovery. As an example, the dynamic stochastic model of a general equilibrium of the Czech real economy is used.

In the first part of the paper a model with the constant parameters is estimated. We use historic shock decomposition of the main macroeconomic variables to identify shocks, which played the most significant role in the recession.

In the subsequent analysis we use the same model, but we relax the typical assumption about structural parameters and we let them to vary in time. Their changes would imply, that the global recession had an influence on the structure of the economy. The economy is represented by the second order approximation of a non–linear DSGE model. The unscented particle filter is used to identify the trajectories of the time–varying parameters.

Keywords: small open economy, shock decomposition, nonlinear DSGE model, time–varying parameters, unscented particle filter.

JEL classification: C32, E32
AMS classification: 91B51, 91B64

1 Introduction

Some authors of economic literature like Fernández-Villaverde and Rubio-Ramírez[3] bring new approach to DSGE models. They think about possible changes in structure of DSGE models caused by changes in economic environment and institutions. Monetary policy and Taylor rule was shown as an example. Authors used second order approximation of DSGE model and shown an algorithm for calculation of this approximation. Also, the non–linear filtering – Particle filter was implemented into estimation.

This approach is used in model for the Czech economy in Tonner and Vašíček [4]. Authors stress transformation and development of the Czech economy during 90s and they study possible structural changes during 2009 decline. Their results show, that some deep parameters changed significantly. The period of economic and financial crisis was analysed in Tvrz and Vašíček [5]. The model of small open economy with financial accelerator used in this paper considers all parameters as time–varying.

The methods of non–linear filtering, especially the Particle filter are described in Andreasen [2]. The general form of PF is derived and consequently modified and optimized to obtain better estimation. The principles and algorithm of the Unscented particle filter are described in Van Der Merwe *et al.* [6].

2 The Model

We use DSGE model of a small open economy similar as in Alpanda *et al.* [1]. The model contains households, intermediate producers, final goods producers, retailers, monetary authority and foreign sector.

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, 324246@mail.muni.cz
[2]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, osvald@econ.muni.cz

Monopolisticaly competitive households supply their differentiated labour on the labour market to intermediate producers. Their wage evolves following the wage Phillips curve. They consume products from final goods producers and imported goods from retailers.

The structure of the sector of firms consists of three stages of production. The intermediate goods producers use labour as the only production factor and produce differentiated goods, which are supplied to final goods producers. Final goods producers use intermediate goods as their factors of production and supply their production to households. Retailers buy foreign goods and sell it to domestic households.

The central bank sets the value of nominal interest rate following the Taylor rule. Foreign economy is described as a closed version of domestic economy.

3 Estimation

3.1 Time–varying parameters

Each parameter θ which is allowed to vary in time is described as unobserved endogenous variable with law of motion

$$\theta_t = (1 - \alpha_t^\theta)\theta_{t-1} + \alpha_t^\theta \bar{\theta} + \nu_t^\theta,$$

where $\bar{\theta}$ is the initial value of the parameter, α_t^θ is an adhesion parameter and $\nu_t^\theta \sim N(0, \sigma_t^\theta)$ is the exogenous shock on θ_t. The parameter α_t^θ influences an ability of the time–varying parameter to return to its initial value. For $\alpha_t^\theta = 0$, the time–varying parameter is defined as a random walk, while the value $\alpha_t^\theta = 1$ implies θ as a white noise centered around its initial value. Adhesion parameter is estimated as time–varying with initial value 0.25 and adhesion 0.25. The standard deviation of an error term ν_t^θ is set on the value $c \cdot \epsilon_\theta$, where c is positive constant and ϵ_θ is a standard error of the estimated constant parameter θ. The constant c is set on the value 4.5.

The first step of the estimation is the application of the Kalman filter on the first-order approximated model. The model with constant parameters is estimated to obtain standard deviations and autoregression parameters of constant structural shocks. Then, the posterior means of parameters are used as an initial values for the time–varying parameters.

3.2 Second order approximation

The Particle filter is applied on the second–order approximated model. The second–order approximation of the law of motion for the endogenous variables can be expressed as

$$\bar{S}_{t+1} = \Psi_{S1}(\bar{S}_t', \epsilon_t') + \frac{1}{2}(\bar{S}_t', \epsilon_t')\Psi_{S2}(\bar{S}_t', \epsilon_t')' + \Psi_{S3},$$

where $\bar{S}_t$ denotes set of the transformed endogenous variables including the time varying parameters, Ψ_{S1} and Ψ_{S3} are vectors and Ψ_{S2} is matrix. The term $\Psi_{S1}(\bar{S}_t', \epsilon_t')$ is the linear solution of the model, $\frac{1}{2}(S_t', \epsilon_t')\Psi_{S2}(\bar{S}_t', \epsilon_t')'$ denotes the quadratic component and Ψ_{S3} is a vector of constants.

The second perturbation of the law of motion for the set of observables Y_t is very similar

$$Y_t = \Psi_{O1}(S_t', \epsilon_t') + \frac{1}{2}(S_t', \epsilon_t')\Psi_{O2}(S_t', \epsilon_t')' + \Psi_{O3}.$$

Again, Ψ_{O1} and Ψ_{O3} are vectors and the term Ψ_{O2} denotes constant matrix. The states in time t and $t-1$ are aggregated in term S_t.

3.3 Unscented Particle Filter

The standard Kalman filter is suitable for the linear state space systems with Gaussian noise. In case of non–linear state space model the Extended Kalman filter(EKF) can be used, nevertheless it is unstable and performs poorly. Particle filtering enables to estimate states of non–linear systems with non–Gaussian noise. For our purposes, Unscented particle filter is used in estimation. It combines Unscented Kalman filtering and Particle filter.

The algorithm of UPF can be briefly described as follows:

1. Initialization: In period $t = 0$, set the prior mean $\bar{x}_0$ and covariance matrix $\mathbf{P}_0$ for the state vector $\mathbf{x}_t$.
2. Generating particles: The mean $\bar{x}_t$ and covariance matrix $\mathbf{P_t}$ are used for generating the total N particles $x_t^{(i)}$, $i = 1, \ldots, N$ from the distribution $p(\mathbf{x}_t)$.
3. Time Update: Use the transition and measurement equation and sigma points and calculate means $\bar{x}_{t|t-1}^{(i)}$, $\bar{y}_{t|t-1}^{(i)}$ and covariance matrices $\mathbf{P}_{t|t-1}^{(i)}$, $\mathbf{P}_{y,y}^{(i)}$ and $\mathbf{P}_{x,y}^{(i)}$.
4. Unscented Kalman filter: Kalman gain $K_t^{(i)} = \mathbf{P}_{x,y}(\mathbf{P}_{y,y})^{-1}$ for each particle $x_t^{(i)}$, $(i = 1, \ldots, N)$ is calculated. Then, $\bar{x}_t^{(i)} = \bar{x}_{t|t-1}^{(i)} + K_t^{(i)}(y_t - \bar{y}_{t|t-1}^{(i)})$ and $P_t^{(i)} = \mathbf{P}_{t|t-1}^{(i)} - K_t^{(i)}\mathbf{P}_{y,y}^{(i)}(K_t^{(i)})^T$.
5. Algorithm continue by the first step, until $t = t_{max}$

In our estimation, 20 runs of the UPF with 30000 particles were performed.

3.4 Data

For the purposes of estimation quarterly time series of seven observables from 1996Q2 to 2013Q4 were used. The domestic economy was described by real GDP, Consumer Price Index and nominal interest rate (PRIBOR). Foreign economy is represented by EA (12) countries, therefore real GDP and Consumer Price Index of Euro area and nominal interest rate (EURIBOR) are used. Connection between domestic and foreign economy is represented by CZK/EUR nominal exchange rate. All series are logarithmed and HP detrended. Also the first difference of both domestic and foreign CPI and exchange rate were taken to obtain CPI inflation deviation from HP trend and exchange rate depreciation.

4 Results

We identify two events accompanied by significant changes in parameters. The first one is the monetary crisis in the second quarter of 1997 and consequent recession during 1998 and 1999. This is closely linked to the change of exchange rate regime in 1997. In the first quarter of 1996 the fluctuation range of fixed exchange rate was changed from 5% to 15%. In May 1997, central bank changed fixed regime to managed floating.

The second economic event, which is more significant for this analysis is the financial crisis of 2007 – 2008, which led to global recession. Start of recession in the Czech economy was delayed compared to our main export destination Germany (the second quarter of 2008). Data from the Czech statistical office show the economic decline in the fourth quarter of 2008. Even though the recession in most countries ended by 2009 and new period of economic growth started again, Czech economy is stagnating.

4.1 Model with constant parameters

The results of the Bayesian estimation are written in Table 1. Only parameters, that are estimated as time varying, are written there. Random Walk Metropolis Hastings algorithm was used for the estimation. The posterior values of the model parameters estimation were used as initial values for time–varying parameters.

Decomposition of output is depicted in Figure 1. Demand, productivity and cost–push shock played significant role during economic growth period of 2006–2008. The last four quarters before economic downturn of 2009 were characterized by significant increase of risk premium for domestic economy. Subsequent decrease of GDP was caused by negative demand shock, which persisted through three quarters, then cost–push shock became important in persistent recession. Monetary shock played role in the first three quarters of 2009 as well. Positive demand shock during the second half of 2010 caused return of the output to steady state, nevertheless during 2012 demand shock together with cost–push shock cut down the output again. Negative contribution of demand shock weakened during 2013 and the wage and productivity shocks had the most significant influence. In the last quarter of 2013 positive contribution of demand shock caused growth of output above long–term trend. This increase of demand is probably linked to unconventional policy of central bank and devaluation of CZK.

Parameter	Description	Distribution	Prior	Prior std.	Posterior	Conf. interval	
ϑ	habit in consumption	B	0.700	0.1000	0.6859	0.5998	0.7733
γ	share of import	B	0.200	0.0250	0.1813	0.1433	0.2189
φ	elasticity. in labour supply	N	3.000	0.6000	3.0288	2.0494	3.9898
ξ_w	elasticity. in labour demand	Γ	1.000	0.4000	0.9927	0.3875	1.5930
δ	price indexation	B	0.250	0.1000	0.2049	0.1316	0.2766
η	el. in substitution dom./for.	Γ	1.000	0.2000	0.6650	0.5141	0.8117
α	wage indexation	B	0.750	0.1000	0.6816	0.4980	0.8731
χ	debt elasticity	B	0.010	0.0025	0.0093	0.0054	0.0131
σ	risk aversion	N	1.000	0.2000	0.8568	0.5507	1.1593
θ_w	Calvo wages	B	0.500	0.1000	0.5103	0.3411	0.6802
θ_h	Calvo domestic	B	0.750	0.1000	0.4225	0.2883	0.5606
θ_f	Calvo importers	B	0.750	0.1500	0.6535	0.5115	0.8008
ρ_r	smoothing parameter	B	0.730	0.1500	0.8052	0.7580	0.8547
ϕ_π	Taylor: inflation	Γ	1.500	0.1250	1.5612	1.3584	1.7555
ϕ_y	Taylor: output	Γ	0.500	0.1250	0.5489	0.3405	0.7510

Table 1 Selected parameters.

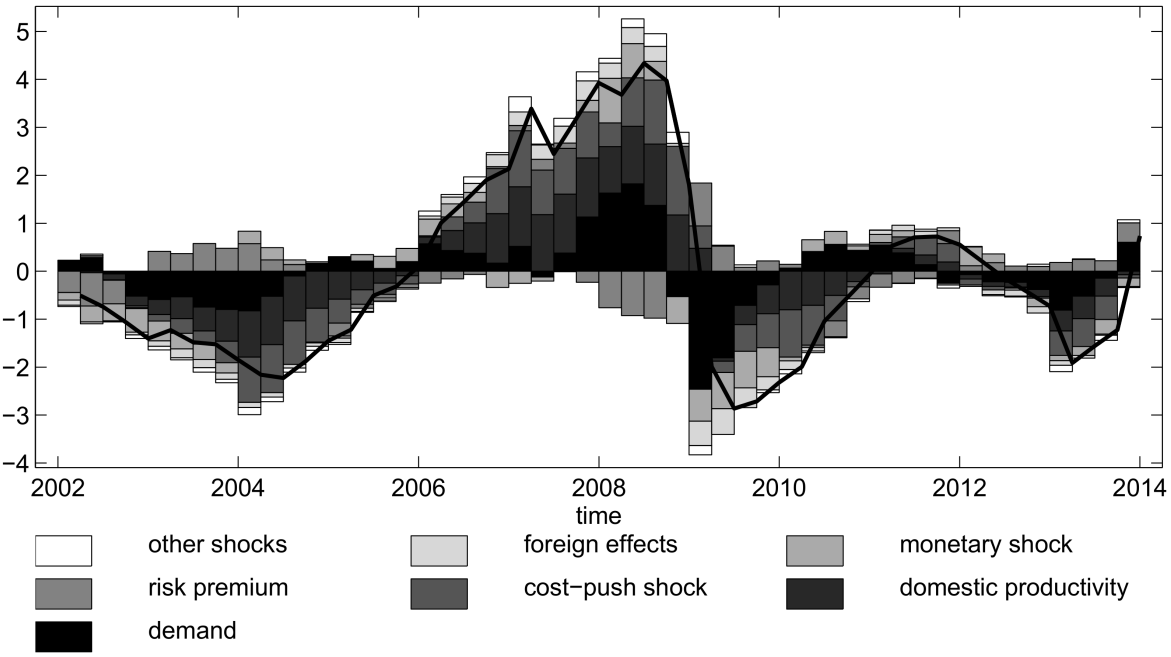


Figure 1 Shock decomposition of output.

4.2 Time–varying parameters

Parameter, which is not biased from its initial value estimated in constant model can be considered as non–varying. There are some parameters, which are stable over time. They are price indexation δ, wage indexation α, Calvo parameter for imported goods θ_f, inverse elasticity of labour supply ϕ, risk premium debt elasticity χ, smoothing parameter in Taylor rule ρ_r, shock gain λ and adhesion parameter.

Parameters, considered as time–varying are depicted in Figure 2. Risk aversion parameter σ, or also inverse elasticity of intertemporal substitution changed significantly during 2008. The decrease of this parameter can be observed from 2006. Households increase their willingness to substitute between present and future consumption, i. e. they consume more and borrow. Incoming recession changed their behaviour and risk aversion increased. Consequences can be seen in the change of the consumption habit parameter ν. Households increased savings and decreased their usual consumption. Significant reduction

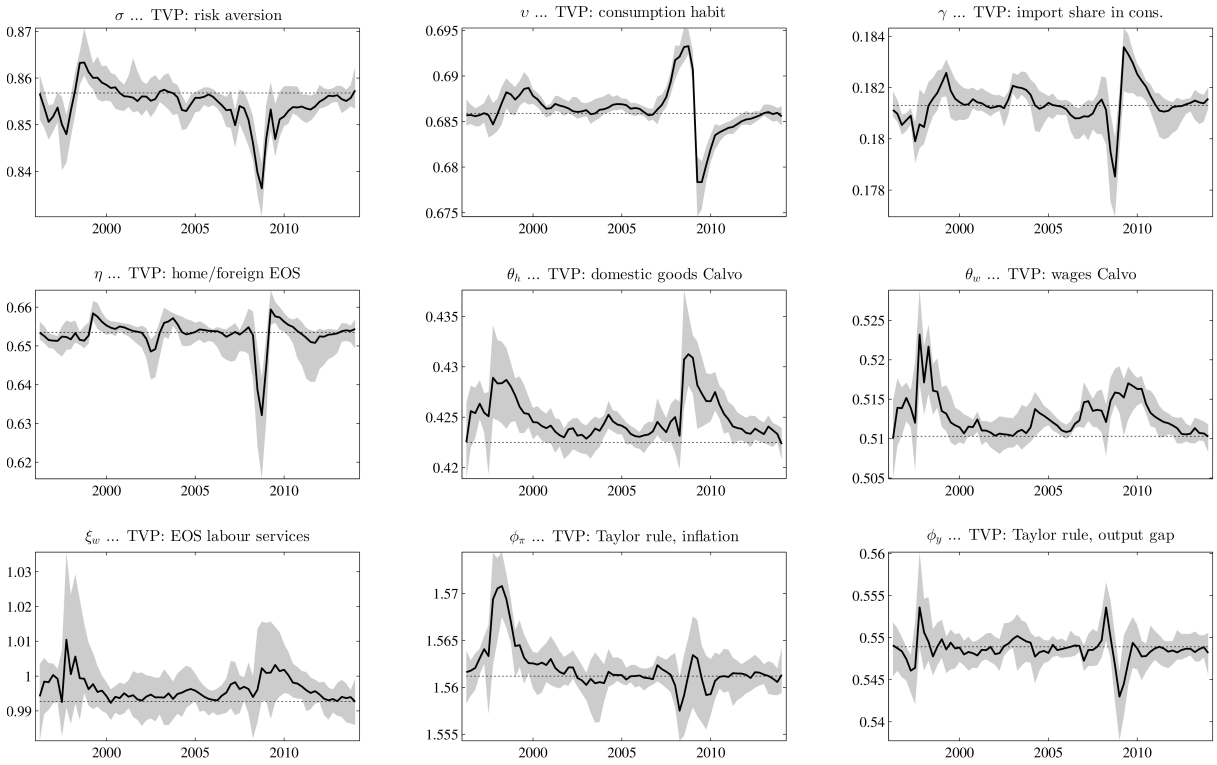


Figure 2 Time–varying parameters.

occurs in the beginning of 2009. Both risk aversion and habit in consumption return back to their initial values quite slowly. These changes contribute to significant decrease of demand during the recession.

Share of import in consumption γ and the elasticity of substitution between home and foreign production consumption η move in the same direction. The trajectory of both γ and η is linked to real exchange rate. Depreciation of exchange rate decreases the relative price of exports, therefore the imports become more expensive. That should decrease the amount of consumed imported goods, nevertheless depicted paths of both γ and η show opposite effect. This can be explained as imperfection of model, which adjusts trajectories of parameters to adapt to the data. Possible economic explanation could be that domestic importers set their prices following their price index. Calvo parameter ϑ_f for importers is quite high. During 2007 to 2008 the real exchange rate appreciated, so the foreign production became relatively cheaper. Importers refused to react by lowering the prices, therefore consumers cut down their demand for foreign goods. Prices of imports decreased in the third quarter of 2008 and share of imports in consumption and elasticity of substitution increased. Recession striking the economy in the fourth quarter of 2008 led to depreciation of exchange rate. Reaction of parameters γ and η is faster, the decrease of them started in the second quarter of 2009.

Calvo parameter for domestic goods producers θ_h heightened during 2008, which implies higher rigidity in prices. Expected recession caused decrease of the demand for goods and consequently decrease of the price level. Producers weren't able or willing to lower their prices, therefore prices stayed on the constant level.

Calvo parameter for labour θ_w was more volatile and grew from 2004 to 2010, there are peaks in 2004, 2007 and the last one in 2009. Low demand for goods caused lower demand for labour. Nevertheless, workers were not willing to decrease their wage, so rigidity grew. Elasticity of labour demand has very similar path. In 2009, producers were most sensitive to price of labour. This led to decrease of productivity. Shock decomposition of output gap shows significant negative contribution of productivity shock influencing the output gap during 2009.

Parameters in Taylor rule are considered as time–varying as well. Monetary authority use Taylor rule to influence output gap and inflation rate. Changes of parameters are linked to changes in the targets of central bank. In case of Czech National Bank, the weight on the inflation rate ϕ_π is approximately three times higher than weight on the output gap ϕ_y. In 2008, the weight on inflation rate decreased

and the weight on output increased. That was a reaction on a strong economic growth, CNB tried to slow down an overheating economy. In the beginning of 2009, weight on inflation grew and weight on output decreased. That is quite unintuitive, because this change implies more restrictive policy during the recession. Possible explanation of this can be a fiscal policy, which is not included in the model. Fiscal consolidation during 2009 influenced output negatively, nevertheless central bank reaction was weak. Model explained this situation as change in central bank policy and parameters changed.

5 Conclusion

The goal of this paper was to identify and analyse possible changes in the structure of the Czech economy represented by the dynamic stochastic model of general equilibrium. The methods of non–linear filtering were used to obtain trajectories of parameters.

The recession, which began in the fourth quarter of 2008, was mostly driven by shocks in domestic demand, productivity and by cost–push shock. Parameters, which were considered as time varying, were linked to behaviour of domestic consumers, rigidity in prices of goods and wages and to changes in the policy of monetary authority.

There is some evidence that changes in the monetary authority policy as well as changes in rigidity of wages may have contributed to deeper economic downturn during 2009. The period of 2010–2013 was characterized by the weak initial economic growth and subsequent second period of recession. Changes in consumers' behaviour may contribute to stronger decrease of domestic demand during the recession and to weak demand in the subsequent years.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0808/2013.

References

[1] Alpanda, S., Kotzé K. and Woglom, G. *The Role Of The Exchange Rate In A New Keynesian Dsge Model For The South African Economy.* South African Journal of Economics, Economic Society of South Africa, vol. 78(2), pages 170-191, 06, 2010.

[2] Andreasen, M. M. *Non-linear DSGE Models and The Optimized Particle Filter.* CREATES Research Paper, 2010-5.

[3] Fernández-Villaverde, J. and Rubio-Ramírez, J. F. *How Structural Are Structural Parameters.* NBER Working Paper 13166, 2007.

[4] Tonner, J. and Vašíček, O. *Parameter Drifting in the Second Order Approximated Model.* Brno: Centrum výzkumu konkurenční schopnosti česk ekonomiky, 2011. 37 s. CVKSCE Working Papers, 4/2011. ISSN 1801-4496

[5] Tvrz, S and Vašíček, O. *Nonlinear DSGE model of the Czech economy with time-varying parameters.* In Vojackova, H. Proceedings of the 31st International Conference Mathematical Methods in Economics. 1. vyd. Jihlava: College of Polytechnics Jihlava, 2013. s. 968-973, 6 s. ISBN 978-80-87035-76-4.

[6] Van Der Merwe, R., Doucet, A., De Freitas, N. and Wan, E. *The Unscented Particle Filter.* CUED Technical Report 380, Cambridge University Engineering Department, 2000.

Possibilities of using Markov chains for construction of technical analysis indicators of stock exchange index PX

Svoboda Milan[1], Lukáš Ladislav[2]

Abstract. The paper brings a continuation of our research topic, which is focused on usage of Markov chains for construction of technical analysis indicators, in general. Our previous results have already confirmed an idea that definition of MC state space based upon accumulated change of underlying asset during period of local monotone trend seems to be rather promising. Here, we present three models of construction of technical analysis indicator, which differ each other with their state spaces being defined in accordance with the accumulated local stock exchange changes. On these models, we define different trading strategies using relatively simple issue – a selected state generates a buying signal, whereas another one generates a selling one, on the contrary. Hence, we get various sequences of buying and selling signals on the MC state spaces. Profitability of such trading strategies is compared with closing-day market values of the PX stock exchange index during the period of nine years, from 2004Q1 till 2012Q4, in particular. Comparing calculated yields following our selected trading strategies with actual PX index values enables us to detect their performances. Except accumulated yields during whole period of nine years we are interested in current changes of invested capital, too, and we see that in some cases we yet outperformed the PX index.

Keywords Markov chain analysis, transition probability matrix, stock exchange index, technical analysis indicators, trading strategies.

JEL Classification: C02, C13, G14, G19
AMS Classification: 90C40, 91B82

1 Introduction

The present study is focused on implementation of Markov chain theory for construction of technical analysis (TA) indicators. TA represents one of prospective possibilities used for prediction of future stock prices. Under TA we understand a large set of methods and instruments which try to predict future prices using past ones and corresponding trade volumes. Academicians, on opposite to traders, were rather diffident ones as far as TA. The first reason thereon stem from generally accepted Efficient Market Hypothesis, see Fama [3], which eliminates possibilities to get superior profits than average ones theoretically. The second reason contributing the first one has appeared due to several early and widely cited negative empirical studies of technical analysis in the stock market, such as Fama and Blume [4], and Jensen and Benington [5]. However, further studies such as Sweeney [8] and Brock et al.[1] have shown that TA methods are capable of outperforming the market, too. Now in general, because of sinking costs of computing power and development of electronic databases one can observe increasing number of works studying profitability and financial efficiency of various TA methods.

Markov chain (MC) is a sequence of discrete random variables X_1, X_2, X_3, ..., obeying Markov property as follows:

$$P\left(X_{n+1} = x_{n+1} \mid X_1 = x_1, X_2 = x_2,.., X_n = x_n\right) = P\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right).$$

In words, MC is a stochastic process defined on discrete set of instants and having discrete set of states that probability $p_i(n)$, expressing a chance the process arrives in state i at instant t_n, depends on state visited in previous instant t_{n-1}, only. All realizations of X_i denoted x_i belong to countable set S = $\{s_i\}$, i = 1, 2,...., N, which is called a state space. Behavior of such process is given by:

- vector of unconditional probabilities $\boldsymbol{p}(n) = (p_1(n), p_2(n), ..., p_N(n))^T$, n = 0, 1, 2, ... where ^{T} denotes transposition,
- transition probability matrix $\boldsymbol{P}$ which elements give conditional probabilities $p_{ij} = P(X_n=s_j \mid X_{n-1}=s_i)$, i,j = 1, 2, .., N, where p_{ij} may depend on n, in general. In case that p_{ij} does not depend on n the MC is called homogeneous one, on opposite to non-homogeneous one.

[1] University of West Bohemia, Husova 11, 306 14 Pilsen, Czech Republic, e-mail: svobodm@kem.zcu.cz.
[2] University of West Bohemia, Husova 11, 306 14 Pilsen, Czech Republic, e-mail: lukasl@kem.zcu.cz.

An idea to apply MC analysis to stock market is not new. From recent papers, we may refer to Doubleday and Esunge [2], Vasanthi et al. [9], or Zhang and Zhang [10]. However, state spaces used within these papers are relatively simple ones. Another way to definition of a state space is given by Svoboda and Lukáš [6] having been presented already on MME 2012. That paper brought four different models, and quite promising results were provided by the fourth one, in particular. The model was defined on filtered sequence of accumulated yield changes of the Prague stock exchange index PX. Now, following our approach we present three new models each one with several trading strategies. Computed results are compared and checked by actual PX index.

2 General definition of trading model

Profitability of any trading strategy will be checked by day values of the PX index within the period 2004 – 2012 (from Jan. 5, 2004 till Dec. 28, 2012, precisely). Well, the day closing values serve us for construction of our TA indicator K_t expressing cumulated yields on local base periods, which are defined in particular by basis index of day closing prices having a local base period which starts to be registered by any trend reverse date in general, i.e. either for monotone increasing or decreasing local sequence of prices. The indicator K_t is defined as follows

$$K_t = K_{t-1}\frac{P_t}{P_{t-1}}, \quad \text{if } P_{t-2} \le P_{t-1} \le P_t, \text{ or } P_{t-2} \ge P_{t-1} \ge P_t \text{ holds}$$

$$K_t = \frac{P_t}{P_{t-1}}, \qquad \text{otherwise}$$

P_t is closing PX value on day t,

P_{t-1} is closing PX value on day t–1,

P_{t-2} is closing PX value on day t–2.

The calculation of K_t is illustrated on Tab. 1.

t	1	2	3	4	5	6	7	8	9
P_t	994,3	1003,3	992,1	975,6	962,6	983,2	956,9	963,1	995,1
K_t		1,0091	0,9888	0,9724	0,9594	1,0214	0,9733	1,0065	1,0399

Table 1 own calculation

Using the calculated values of K_t we may define various state spaces of our models. Within each model we also define several trading strategies. As for any strategy itself, we follow a simple idea that a specified state belonging to the state space of particular model issues a buying signal, whereas the other one issues a selling signal.

For all models we apply a unified checking procedure to test their profitability. The procedure is based on transactions, which represents trades. A transaction represents one buying an asset and its subsequent selling. If a buying or selling signal is issued one day, the trading price of the trade is defined by opening price next day. We accept trading with whole capital only, thus eliminating any fractional investments. Further, any transaction costs are neglected, and we also prevent allowing both short selling and two consecutive buying, as well. The value of invested capital is calculated by formula:

$$C_i = C_{i-1}\frac{S_i}{B_i},$$

which gives a formula for accumulated value after n trades realized as follows

$$C_n = C_0 \prod_{i=1}^{n} \frac{S_i}{B_i}, \text{ where}$$

C_0 = 1.00 [m.u.] is an initial investment capital in a monetary unit,

C_n gives the value of invested capital after n trading transactions,

S_i represents selling price for i-th trading transaction,

B_i represents buying price for i-th trading transaction.

3 Models

3.1 Model 1

Using K_t, we define the state space $S = \{s_i\}$, i=1,2,..,8 containing eight states. The four states denoted {D_4, D_3, D_2, D_1} are introduced to capture different levels of decreasing value, where D_4 catches the deepest decline and the D_1 the smallest one. In similar way, the four states denoted {G1, G2, G3, G4} are introduced to capture increasing values, and G_1 catches the smallest growth and the G_4 the biggest one. Quantified levels of K_t with corresponding states assigned are as follows

D_4: $K_t < 0{,}97$; D_3: $0{,}97 \le K_t < 0{,}98$; D_2: $0{,}98 \le K_t < 0{,}99$; D_1: $0{,}99 \le K_t < 1{,}00$;

G_1: $1{,}00 \le K_t < 1{,}01$; G_2: $1{,}01 \le K_t < 1{,}02$; G_3: $1{,}02 \le K_t < 1{,}03$; G_4: $1{,}03 \le K_t$.

Now, we are going to define trading strategies on the state space created. First, we define 16 the simplest strategies. Each one is formulated by a couple (D_i, G_j), i,j=1,..,4 where state D_i generates the buying signal and G_j generates the corresponding selling signal. We have calculated accumulated value C_n, together with the corresponding number of transactions n realized, for all 16 trading strategies introduced. All results are presented on the Tab. 2.

sell / *buy*	G_1		G_2		G_3		G_4	
	C_n	n	C_n	n	C_n	n	C_n	n
D_1	1,2730	290	0,5958	211	1,2341	129	2,1265	106
D_2	1,6170	164	1,6185	140	2,3181	104	2,3669	90
D_3	0,5976	87	0,3832	81	0,6527	72	0,7993	65
D_4	0,5976	74	0,4789	67	0,9127	66	1,1298	72

Table 2 – own calculation

In addition to these global results over whole nine years long period investigated, we are concern to follow time dependent development of capital invested value over the whole period, as well. For illustration purposes, we select just three trading strategies from all 16 defined ones, and the results are presented on the Fig. 1.

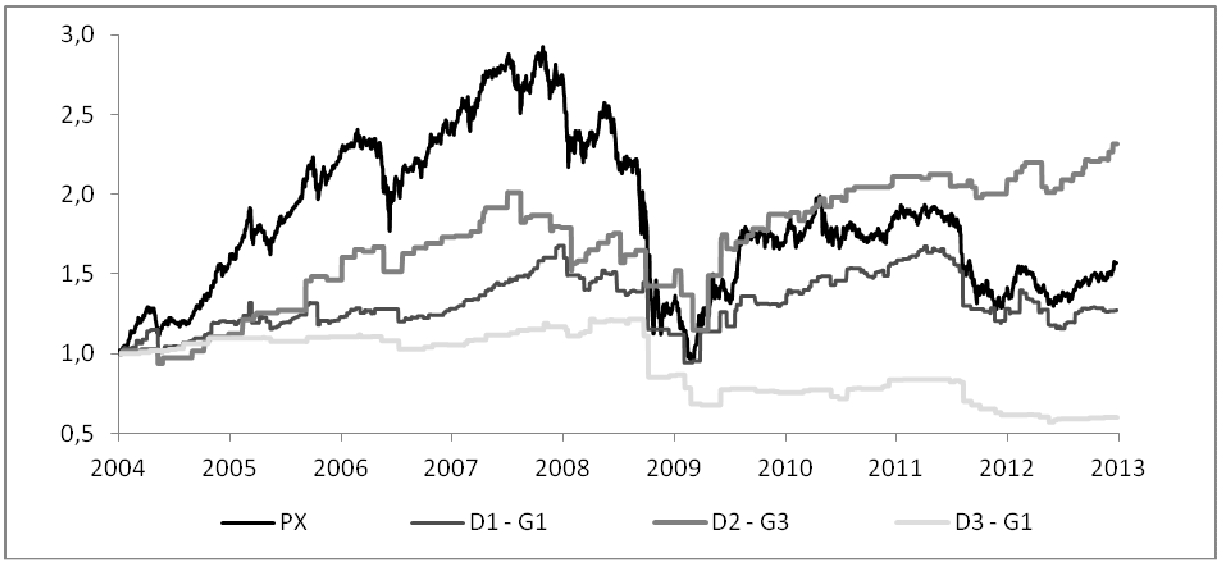


Figure 1 – own calculation

Checking profitability of all our 16 trading strategies by the earnings ratio of the PX index 1.5689 on the time period investigated, we may conclude that in five cases we have outperformed it. The two highest profitability results are gained by strategy (D_2,G_3) and (D_2,G_4) when inspecting the results presented in the Tab.1. The Fig. 1 also shows that any of our trading strategies did not grasp a period of progressive PX profitability increase during period, from 2004 till the first half of 2007. On the contrary as a positive issue, we may note that any of them did not sink relatively as much as the PX did within the next period, i.e. from the second half of 2007 till the beginning of 2009.

3.2 Model 2

The second model is created as a derivative of the Model 1 with its profit promising state D_2. We already know that D_2 being selected as buying signal issuing state leads to the best profitability. Hence, we try to investigate if small perturbations of D_2 state definition will lead to even better results prospectively. We will shift a lower bound of D_2 state interval by 0.2% increments while still preserving its width to be 1% precisely (note: we have also investigated the different widths but without any significant differences).

The model consists of seven states denoted B_i, i=1,..,7 and devoted to buying signal issues, and five states denoted S_j, j=1,..,5 which serve as selling signal transmitters. First, the states B_i are defined as follows

B_1: $0{,}984 \le K_t < 0{,}994$; B_2: $0{,}982 \le K_t < 0{,}992$; B_3: $0{,}980 \le K_t < 0{,}990$; B_4: $0{,}978 \le K_t < 0{,}988$;

B_5: $0{,}976 \le K_t < 0{,}986$; B_6: $0{,}974 \le K_t < 0{,}984$; B_7: $0{,}972 \le K_t < 0{,}982$.

Second, the states S_j are defined in a slightly different manner, just by partially overlapping semi-closed intervals instead by disjunctive ones only. It means, that corresponding selling signal is issued as it reaches the specified level of accumulated increase on local base period. They are defined as follows.

S_1: $K_t \ge 1{,}0225$; S_2: $K_t \ge 1{,}0275$; S_3: $K_t \ge 1{,}0325$; S_4: $K_t \ge 1{,}0375$; S_5: $K_t \ge 1{,}0425$.

Thus using this model, we define 35 different trading strategies in total. The results of calculated value C_n and the corresponding number of transactions n realized, again, are presented on the Tab. 3.

sell / buy	S_1		S_2		S_3		S_4		S_5	
	C_n	n	C_n	n	C_n	n	C_n	n	C_n	n
B_1	1,927	141	2,349	112	2,524	91	2,608	67	2,293	53
B_2	2,065	129	2,349	105	2,538	85	2,582	64	2,434	51
B_3	1,868	120	2,364	100	2,730	81	2,673	61	2,538	49
B_4	1,229	110	1,495	94	1,647	79	1,681	62	1,490	50
B_5	1,687	106	1,927	90	1,917	76	1,889	59	1,742	48
B_6	1,954	104	2,465	87	2,328	73	2,075	56	1,765	45
B_7	1,166	91	1,283	76	1,169	63	1,147	51	1,106	40

Table 3 – own calculation

For illustration purposes, we select three trading strategies again from all 35 defined ones, and the results of time dependent development of their financial performance are presented on the Fig. 2.

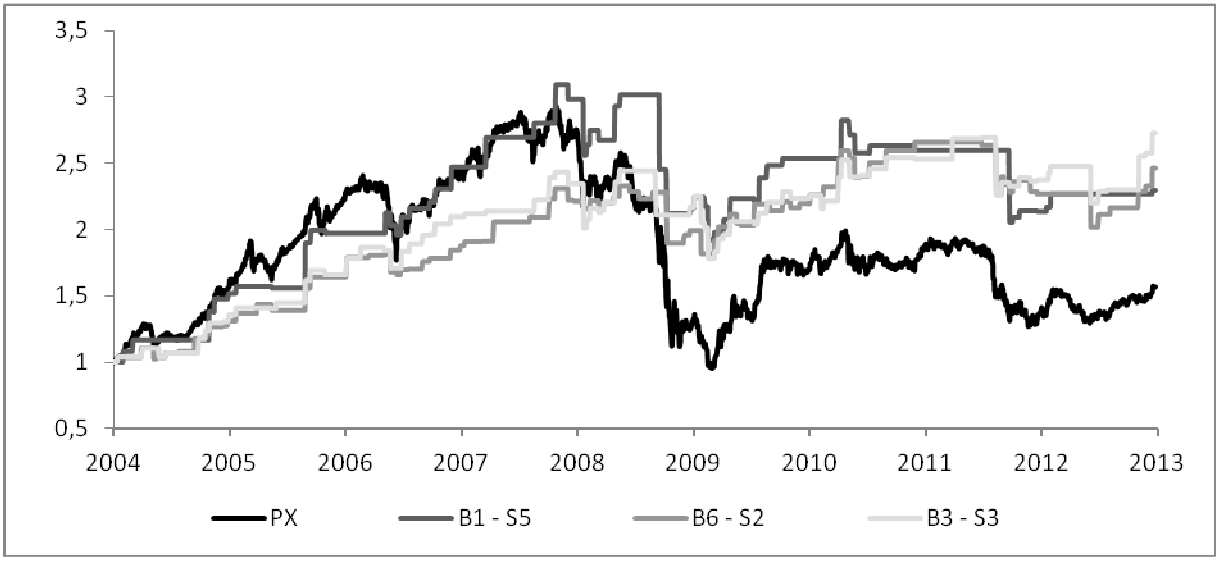


Figure 2 – own calculation

Using the Model 2, we investigated thoroughly all possible 35 trading strategies, and we have got better performance than PX index has, in great majority of cases. The best results have appeared under the strategies with buying signal issues due to entering states B_1, B_2, or B_3, and selling signals issues at S_3, or S_4, respectively. For almost all of these trading strategies we have got even higher results than the maximum from the Model 1 is

(2.3669). In particular, the maximum (2.730) on the Tab. 3 yields the strategy (B_3,S_3). Comparing the Fig. 2 with the Fig. 1 we see that the period of progressive PX profitability increase during period, from 2004 till the first half of 2007, is captured in a better way.

3.3 Model 3

The third model is created as a mollification of the Model 2 in order to prevent prospective sinks of investment capital value. A clue idea is following. If a buying signal is issued at corresponding state but with decreasing trend and the current value is still advancing to drop then we are going to leave the position, i.e. to sell the asset unless its value stands within given drop tolerance as soon as possible. In other words, it means that the selling signal may be issued either by specified level of value increase, or a given level of decrease, as well.

The model consists of three most promising states $\{B_1,B_2,B_3\}$ devoted to buying signal issues identified in the Model2, five states $\{S_1,S_2,S_3,S_4,S_5\}$ of the same model being devoted to issue selling signals in case of value increases specified, and finally, nine particular states issuing selling signals always provided the value drops when exceeding given drop tolerance limits. Such drop tolerance limits are defined as accumulated value lost by specified quantity k. Hence, the selling signal, under the period of local value decrease, will be issued if $K_t \leq d_i - k$. Let assume for example the buying signal is issued by B_1 (B_1: $0{,}984 \leq K_t < 0{,}994$), where it is d_1=0,984. Further, we accept three different drop tolerances k = 0.009, 0.012, and 0.015, in particular. Hence, we introduce three selling signal issue cases: i) $K_t \leq 0.975$ (k=0.009), ii) $K_t \leq 0.972$ (k=0.012), iii) $K_t \leq 0{,}969$ (k=0,012), respectively. All results we calculated are summarized in the Tab. 4.

buy	*sell*	S_1		S_2		S_3		S_4		S_5	
		C_n	n	C_n	n	C_n	n	C_n	n	C_n	n
B_1	k = 0,009	2,314	176	2,626	158	2,756	143	2,615	125	2,753	114
B_1	k = 0,012	2,874	170	3,185	148	3,340	133	3,135	114	3,341	103
B_1	k = 0,015	2,345	164	2,696	141	2,804	124	2,667	105	2,932	94
B_2	k = 0,009	2,870	156	3,052	143	3,106	129	2,981	114	3,485	106
B_2	k = 0,012	2,531	152	2,742	136	2,899	122	2,778	107	3,238	98
B_2	k = 0,015	2,515	147	2,712	130	2,699	113	2,617	96	3,050	87
B_3	k = 0,009	2,137	153	2,713	140	2,936	124	3,047	101	3,318	92
B_3	k = 0,012	2,425	136	2,591	120	3,134	107	2,838	93	3,091	84
B_3	k = 0,015	2,424	130	2,581	113	3,122	99	2,918	84	3,178	75

Table 4 – own calculation

Again as in the previous model, we plot just three trading strategies representing a sample of our numerical calculations. They are depicted on the Fig. 3.

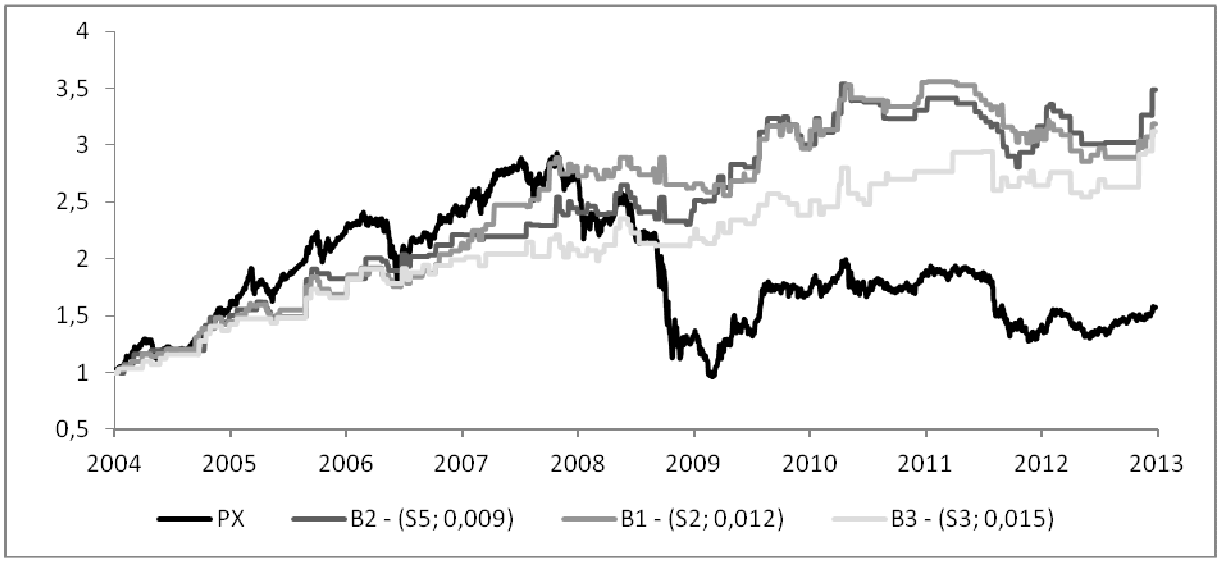


Figure 3 – own calculation

Inspecting the results gained by Model 3 where selling signals are issued also when given drop tolerances are met, we may conclude that the financial performance has improved significantly. Many of applied trading strategies within this model yield better profitability as compared with those ones gained by the Model 2. The global maximum (3.485) over all strategies applied yields the strategy (B_2; $B_2(k=0.009)$,S_3), as given on the Tab. 4. The Fig. 3 shows that we have almost eliminated excessive value drops, too. In particular, the investment capital value yielded by our trading strategies do not nearly drop during heavy drawdown of the PX index value within the period, 2004Q2 – 2009Q1. However, we do not eliminated value drops during 2011.

4 Conclusion

Using our three models for TA indicators based upon Markov chains and applied to Prague stock exchange index PX analysis, we have shown mostly outperforming behavior of our trading strategies proposed over actual PX gains. Ongoing research will be focused on three challenging topics:

- analysis of transaction costs and their influence upon trading strategy profitability being checked by actual PX index gains again;
- investigation of dynamic adaptive definition of state space and corresponding buying/selling signals conditional issuing, since as known financial time series exhibit time dependent volatility thus requesting non-homogeneous MC tools, see e.g. Svoboda and Lukáš [7];
- investigation of short positions, i.e. allowance of trading drop speculative behavior.

Acknowledgements

The research project was supported by Grant SGS-2014-047 of the Faculty of Economics, University of West Bohemia in Pilsen.

References

[1] Brock, W., Lakonishok, J. and LeBaron, B.: Simple Technical Trading Rules and The Stochastic Properties of Stock Returns. *Journal of Finance* **47** (1992), No. 5, 1731–1764.

[2] Doubleday, K., J. and Esunge, J., N.: Application of Markov Chains to Stock Trends. *Journal of Mathematics and Statistics* **7** (2011), (2), 103–106.

[3] Fama, E., F. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance* **25** (1970), No.2, 383–417.

[4] Fama, E., F. and Blume, M., E.: Filter Rules and Stock Market Trading. *Journal of Business* **39** (1966), 283–306.

[5] Jensen, M., C. and Benington, G., A.: Random walks and technical theories: some additional evidence, *Journal of Finance* **25** (1970), 469–482.

[6] Svoboda, M., Lukas, L.: Application of Markov chain analysis to trend prediction of stock indices. In: *Proceedings of* 30*th Int. Conf. Mathematical Methods in Economics* (Ramík, J., and Stavárek, D., eds.). Karviná: Silesian Univ., School of Business Administration, 2012, 848–853.

[7] Svoboda, M., Lukas, L.: Application of non-homogeneous Markov chain analysis to trend prediction of stock indices. In: *Proceedings of* 31*th International Conference Mathematical Methods in Economics* (Vojáčková, H., ed.). Jihlava: College of Polytechnics, 2013, 881–886.

[8] Sweeny, R., J.: Some new filter rule tests: methods and results. *Journal of Financial and Quantitative Analysis* **23** (1988), 285–300.

[9] Vasanthi, S., Subha M., V. and Nambi S., T.: An empirical Study on Stock Index trend prediction using Markov Chain analysis. *JBFSIR* **1**, Issue 1, 72–90.

[10] Zhang, D., Zhang, X.: Study on Forecasting the Stock Market Trend Based on Stochastic Analysis Method. *International Journal of Business and Management* **4** (2009), No. 6, 163–170.

On the use of the balanced half-samples method in the construction of factorial designs

Małgorzata Szerszunowicz [1]

Abstract. Design of experiments, as a tool of statistical quality control, leads to the improvement of technological and economical results of a manufacturing process. In practice, experimental designs which are most frequently used by production companies include full and fractional factorial designs.

The balanced half-samples method is used, inter alia, to estimate the variance of estimators in complex surveys. The present article deals with the use of balanced half-samples method in construction of factorial design of experiments for factors with two levels. The application the idea of balanced half-samples method in the design of experiments allows one to reduce the number of experimental trials and, at the same time, to obtain better results of production processes. The proposed method will be presented for selected empirical data and the results will be compared to classical factorial designs of experiments.

Keywords: design of experiments, balanced half-samples method, response surface function.

JEL classification: C99
AMS classification: 62K20

1 Introduction

Design of experiments is one of the tools of statistical quality control allowing one to effectively improve the results of production processes. As far as classical designs of experiments are concerned, an important role is played by fractional factorial designs of experiments. Their use reduces the number of experimental trials realized during the experiment, which results in a reduction in both the time and cost of the experiment.

The aim of this paper is to construct a factorial design of experiment for factors with two levels which, like fractional designs, leads to a reduction in the number of experimental trials. To this end, the balanced half-sample method will be used, and its results will be presented for some empirical data.

2 The classical factorial designs of experiments

The design of experiments is the most essential phase which precedes production processes, so its application should be carried out in accordance with the following criteria formulated by D. C. Montgomery [4],[5]:

- recognise and define the problem by determining all the aspects. circumstances and potential objectives of the experiment;
- appropriately select the factors, their levels and ranges and explore the possibility of considering them in the experiment;
- define the response variable;
- choose a proper design of experiment, i.e. determine the number of experimental trials and the possible randomization restrictions;

[1]University of Economics in Katowice, Department of Statistics, 1 Maja 50 Street, Katowice, Poland, malgorzata.szerszunowicz@ue.katowice.pl

- perform the experiment;
- analyse the results using statistical methods;
- formulate conclusions and recommendations resulting from the analysis of the results.

The notion of experiment should be understood as a sequence of n experimental trials, where a single experimental trial is a single result of random variable Y, with fixed values of factors $X_1, X_2, \ldots, X_m$. Let us define $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_m}$, as the sets of all the possible values of factors $X_1, X_2, \ldots, X_m$, and then the experimental area is a set of points $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ where $x_i \in \mathbf{X_i}$, for $i = 1, 2, \ldots, m$. The set of pairs $P_n = \{\mathbf{x}_j, p_j\}_{j=1}^n$ defines a design of experiment with n experimental trials, where $\mathbf{x}_j = (x_{1j}, x_{2j}, \ldots, x_{mj})$ and $p_j = \frac{n_j}{n}$, where n_j is the number of experimental trials in point $\mathbf{x}_j$ of the experimental area, and also $\sum_{j=1}^n n_j = n$ and $\sum_{j=1}^n p_j = 1$ for $j = 1, 2, \ldots, n$.

The relationship between the set of random and non-random factors $X_1, X_2, \ldots, X_m$ and the response variable Y can be presented in the form of the following statistical model [7]

$$Y(X_1, X_2, \ldots, X_m) = y(X_1, X_2, \ldots, X_m) + \epsilon \tag{1}$$

where $EY(X_1, X_2, \ldots, X_m) = y(X_1, X_2, \ldots, X_m)$, $E\epsilon = 0$ and $V\epsilon = \sigma^2$, where σ^2 is a constant value, independent of the values of the factors, while the function $y(x_1, x_2, \ldots, x_m)$ is named response surface function. The arguments of the response surface function are m realizations of non-random variables $X_1, X_2, \ldots, X_m$. The model (1) can be presented as a general linear model $\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where

$$\mathbf{Y^T} = (Y_1 \ Y_2 \ \ldots \ Y_n) \tag{2}$$

$$\boldsymbol{\epsilon}^\mathbf{T} = (\epsilon_1 \ \epsilon_2 \ \ldots \ \epsilon_n) \tag{3}$$

$$\boldsymbol{\beta}^\mathbf{T} = (\beta_1 \ \beta_2 \ \ldots \ \beta_k) \tag{4}$$

$$\mathbf{f^T}(\mathbf{x}) = (f_1(\mathbf{x}) \ f_2(\mathbf{x}) \ \ldots \ f_k(\mathbf{x})) \tag{5}$$

$$\mathbf{F} = \begin{bmatrix} f_1(\mathbf{x_1}) & \ldots & f_k(\mathbf{x_1}) \\ \vdots & \ddots & \vdots \\ f_1(\mathbf{x_n}) & \ldots & f_k(\mathbf{x_n}) \end{bmatrix} \tag{6}$$

and $f_i(\mathbf{x}_j) \equiv x_{ij}$, for $i = 1, 2, \ldots, k, j = 1, 2, \ldots, n$. The estimation of the parameters of response surface function $\mathbf{y} = \mathbf{F}\boldsymbol{\beta}$ is carried out with the use of the least squares method.

In the literature ([4], [5], [7]) concerning the classical design of experiments, the following response surface function is considered:

$$y(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \beta_{12} x_1 x_2 + \ldots + \beta_{12\ldots m} x_1 x_2 \ldots x_m. \tag{7}$$

The response surface function (7) is a special case of the form of general linear model given above by (1) - (6), where $\boldsymbol{\beta}^\mathbf{T} = (\beta_0 \beta_1 \beta_2 \ldots \beta_m \beta_{m+1} \ldots \beta_{12\ldots m})$ and $f_1(\mathbf{x}) \equiv 1, f_2(\mathbf{x}) = x_1, f_3(\mathbf{x}) = x_2, \ldots, f_m(\mathbf{x}) = x_{m-1}$, $f_{m+1}(\mathbf{x}) = x_m$, $f_{m+2}(\mathbf{x}) = x_1 x_2, \ldots,$ $f_{2^m}(\mathbf{x}) = x_1 x_2 \ldots x_m$. In justified cases ([7]) it is assumed that the interactions of third and higher factors are equal to 0, and then the response surface function (7) can be expressed with the following equation:

$$y(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \beta_{12} x_1 x_2 + \ldots + \beta_{m-1 \ m} x_{m-1} x_m. \tag{8}$$

The interactions of all the factors in particular may be omitted and then the response surface function is as follows:

$$y(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m. \tag{9}$$

Estimation of the parameters of response surface functions (7), (8), (9) consists in performing an experiment which takes into account n experimental trials, so that the estimation of all the parameters of response surface function is possible. The most frequently used designs of experiments are the 2^m full factorial designs of experiments and 2^{m-p} fractional factorial designs. The 2^m full factorial designs are used to estimate of response surface function (7) with all the interactions between the factors, but this design is characterized by a large number of experimental trials with an increasing number of factors, which makes it difficult to carry out the complete experiment. In order to reduce the number of experimental trials of the performed design of experiment, the 2^{m-p} fractional factorial designs are used. The 2^{m-p} designs for $p = 1$ reduce the number of experimental trials 50%. However these designs do not allow one to estimate any repsonse surface functions, in particular functions (7) [3].

3 The balanced half-samples method

Created in the USA in the 1960s, the balanced half-samples method is used to estimate the variance of parameters in research conducted by statistical institutes (e.g. Eurostat). The balanced half-samples method is applied in stratified sampling for the estimation of the variance of fixed parameter θ, and its scheme can be divided into several stages. The first one ought to create a sample

$$s = \bigcup_{h=1}^{L} s_h \tag{10}$$

consisting of L strata, where each stratum s_h consists of two elements y_{h_1}, y_{h_2}. Next, a half-sample is created by choosing only one element from each stratum s_h, and then each of the L elements of the half-sample is assigned value 1 or values -1 according to the following formula:

$$\varepsilon_{\alpha h} = \begin{cases} 1, & y_{h_1} \in s_h \\ -1, & y_{h_2} \in s_h \end{cases}, \tag{11}$$

where $\alpha = 1, 2, \ldots, k$. A set of k half-samples defined in this way is called a balanced half-samples set, when for every pair of strata $h, h' \in \{1, \ldots, L\}$ the following equality is fulfilled

$$\sum_{\alpha=1}^{k} \varepsilon_{\alpha h} \varepsilon_{\alpha h'} = 0. \tag{12}$$

The set of balanced half-samples allows one to decrease the accounting difficulties which arise when the number of strata is large [2].

4 The method of constructing partial designs of experiments and its application

The classical 2^m full factorial designs of experiments require large numbers of experimental trials, which causes some increase in the costs of experiments and prolongation of the time of their performance. Fractional factorial design reduce both the number of experimental trials and the possibility of selecting appropriate response surface function. In order to estimate the right response surface function and reduce the number of experimental trials, one may use the balanced half-samples method.

4.1 The proposed method of constructing partial factorial designs

Let us assume that the 2^m full factorial design of experiment is given as $P_n = \{\mathbf{x}_j, \frac{1}{n}\}_{j=1}^{n}$, where the m factors occur at two levels: upper $(+1)$ and lower (-1). The designation of factor levels can be equated with the symbols from half-samples according to formula (11). Then the experimental area points $\mathbf{x}_j$ for $j = 1, \ldots, n$ characterize the half-samples, while the values of factors $X_1, X_2, \ldots, X_m$ correspond to the individual strata. Consequently, the construction of partial design of experiment which uses the balanced half-samples method may be presented as follows:

1. determine the full factorial design of experiment P_n with n experimental trials;
2. for every $j = 1, \ldots, n$, according to condition (12), create a set of k balanced half-samples such as $\mathbf{x}_{bhs} = \{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots, \mathbf{x}_{(k)}\}$;
3. eliminate the points of experimental area determined by set $\mathbf{x}_{bhs}$ from the full factorial design of experiment;
4. perform the experiment whose results are the best among all the possible designs.

A design of experiment created in this way includes $n - k$ experimental trials which allow one to estimate the selected response surface function.

4.2 The use of the partial design for selected empirical data

Antony [1] considers an experiment which includes four factors occuring at two levels. This experiment explores the impact of pouring temperature (X_1), titanum content (X_2), heat treatment method (X_3) and amount of grain refiner used (X_4) on the crack length (Y) of an element. The full factorial design P_{16} with the values of response variable Y is presented in Table 1.

No.	X_1	X_2	X_3	X_4	Y
1	-1	-1	-1	-1	7,037
2	1	-1	-1	-1	14,707
3	-1	1	-1	-1	11,635
4	1	1	-1	-1	17,273
5	-1	-1	1	-1	10,403
6	1	-1	1	-1	4,368
7	-1	1	1	-1	9,36
8	1	1	1	-1	13,44
9	-1	-1	-1	1	8,561
10	1	-1	-1	1	16,867
11	-1	1	-1	1	13,876
12	1	1	-1	1	19,824
13	-1	-1	1	1	11,846
14	1	-1	1	1	6,125
15	-1	1	1	1	11,19
16	1	1	1	1	15,653

Table 1 The full factorial design of experiment P_{16}

Let us assume that the purpose of experimenter's research is to estimate the parameters of the following response surface function without factor interactions:

$$y(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \tag{13}$$

and the following response function with selected interactions between the included factors

$$y(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4, \tag{14}$$

on the assumption that the smallest number of experimental trials is used and the response surface function has the best possible match.

Using the classical designs of expeirments in order to estimate response surface parameters in the form of (13) one may define fractional factorial design 2^{4-1} presented in Table 2, which includes half of the experimental trials of design P_{16}.

No	X_1	X_2	X_3	X_4	Y
1	-1	-1	-1	-1	7,037
2	1	1	-1	-1	17,273
3	1	-1	1	-1	4,368
4	-1	1	1	-1	9,36
5	1	-1	-1	1	16,867
6	-1	1	-1	1	13,876
7	-1	-1	1	1	11,846
8	1	1	1	1	15,653

Table 2 The fractional factorial design 2^{4-1}

The fractional factorial design which uses 8 experimental trials allows one to estimate the parameters of response surface function (13), giving the function in the form of

$$y(\mathbf{x}) = 11,61 + 1,08x_1 + 1,58x_2 - 1,3x_3 + 2,95x_4 \tag{15}$$

and the value of coefficient $R^2 = 0,848$, which shows a good match for the experimental data.

In the case of response surface function with all factor interactions (14) the use of fractional design 2^{4-1} is not possible because the number of parameters with the intercept is greater than the number of experimental trials of 2^{4-1} design. Consequently, one may use a partial design of experiment which uses the balanced half-samples method.

According to the scheme presented in chapter 4.1, the full factorial design of experiment has been fixed as in Table 1. For each point of the experimental area the sets of $k = 4$ balanced half-samples have been defined as follows:

$$\mathbf{x}_{bhs}^{(j,i)} = \{\mathbf{x}_{(1)}^i, \mathbf{x}_{(2)}^i, \mathbf{x}_{(3)}^i, \mathbf{x}_{(4)}^i\}, \text{for } j = 1, 2, \ldots, 16,\ i = 1, 2, \ldots, l, \tag{16}$$

where l denotes the number of all the possible sets of balanced half-samples for a fixed point of experimental area $\mathbf{x}_j$. As a result, 32 different sets of balanced half-samples were determined. Next, the full factorial design was reduced by the experimental trials defined by the successive sets of balanced half-samples. Each of the created designs includes 12 experimental trials, which makes it possible to estimate response surface parameters (14). For every estimated response surface function the coefficient R^2 was computed. The greatest value of coefficient $R^2 = 0,99998$, i.e. the best match for the experimental data, was obtained for the following design without experimental trials determined by the set of balanced half-sample $\mathbf{x}_{bhs}^{(3,6)}$ presented in Table 3. The created partial design of experiments is presented in Table 4.

X_1	X_2	X_3	X_4
-1	1	-1	-1
-1	-1	1	-1
-1	-1	-1	1
-1	1	1	1

Table 3 The set of balanced half-sample $\mathbf{x}_{bhs}^{(3,6)}$

No.	X_1	X_2	X_3	X_4	Y
1	-1	-1	-1	-1	7,037
2	1	-1	-1	-1	14,707
3	-1	1	-1	-1	11,635
4	1	-1	1	-1	4,368
5	-1	1	1	-1	9,36
6	1	1	1	-1	13,44
7	1	-1	-1	1	16,867
8	-1	1	-1	1	13,876
9	1	1	-1	1	19,824
10	-1	-1	1	1	11,846
11	1	-1	1	1	6,125
12	1	1	1	1	15,653

Table 4 The partial design of experiment

The estimated response surface function is

$$\begin{aligned} y(\mathbf{x}) &= 12,03 + 1,5x_1 + 2x_2 - 1,73x_3 + 2,53x_4 + 1,01x_1x_2 \\ &- 1,91x_1x_3 - 1,44x_1x_4 + 1,63x_2x_3 + 0,11x_2x_4 - 0,09x_3x_4. \end{aligned} \tag{17}$$

The criteria for the selection of a balanced half-samples set may be different depending on the studied phenomenon. The selection of the set $\mathbf{x}_{bhs}^{(j,i)}$ can be determined by the experimenter, historical results of experiment or the conditions under with the experiment is carried out.

Proposed example of partial design of experiment shows that using the idea of balanced half-sample method allows one to estimate the parameters of response surface function based on only one measurement of response variable. That leads to obtain better results of production processes and to reduce the costs and time of the experiment.

The use of the balanced half-samples method in defining the design of experiments may lead to specifying experimental designs which may be an alternative to 2^{m-p} fractional designs of experiments. In the above calculations for $p = 1$, in particular was obtained the comparison with the fractional factorial design of experiment containing half of experimental trials.

5 Conclusions

The factorial design of experiments plays an important role in planning the results of production processes. Determination of the appropriate response surface function which describes the manufacturing process in the best possible way is an important element of the realization of the design of experiment. The estimation of response surface function requires selection of design of experiment which allows one to estimate all the parameters of a selected function. Full factorial design of experiments makes it possible to estimate all the parameters, but its realization results in additional costs and prolonged time of production processes. On the other hand, the fractional factorial design of experiment does not allow one to estimate response surface with the number of parameters greater than the number of experimental trials. Consequently, partial design using the balanced half-sample method to eliminate experimental trials so that the estimation of response surface is possible may constitute an alternative form of design of experiments.

Acknowledgements

Supported by Polish National Science Center grant No. DEC-2011/03/B/HS4/05630.

References

[1] Antony, J.: *Design of Experiments for Engineers and Scientists*. Butterworth - Heinemann, Oxford, 2003.

[2] Jakubowski, J.: *Metoda półprób zrównoważonych i jej uogólnienia*. Główny Urząd Statystyczny, Warszawa, 1999.

[3] Kończak, G.: *Metody statystyczne w sterowaniu jakością produkcji*. Wydawnictwo Akademii Ekonomicznej, Katowice, 2007.

[4] Montgomery, D. C.: *Design and Analysis of Experiments*. John Wiley & Sons, Inc., New York, 2001.

[5] Montgomery, D. C.: *Introduction to statistical quality control*. John Wiley & Sons, Inc., New York , 1997.

[6] Ryan, T. P.: *Modern Experimental Design*. John Wiley & Sons, New Jersey, 2007.

[7] Wawrzynek, J.: *Planowanie eksperymentów zorientowane na doskonalenie jakości produktu*. Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław, 2009.

Small Open Economy Real Business Cycle Model of the Slovak Economy

Karol Szomolányi, Martin Lukáčik, Adriana Lukáčiková[1]

Abstract. We calibrate a simple small open real business cycle model using the Slovak data. The second moments of the Slovak aggregate cyclical components better fit the theoretical model, if we consider the CES production function.
Since Slovak economy is small open and integrated to the European Monetary Union, in our view terms of trade shocks explain the cyclical properties of the real GDP, consumption, investment, trade balance and current account. Terms of trade shocks have the similar economic short-term impact as the productivity shocks; therefore we consider the small open economy business cycle model as the suitable tool to explain the Slovak business cycles.
Furthermore even if the Slovak economy is rather emerging country, its second moments of the given Slovak aggregate cyclical components better corresponds to the small developed country: consumption is less volatile than output, government purchases are low volatile and countercyclical, trade balance and current account are countercyclical. Therefore shocks significant for the emerging countries business cycles cannot explain Slovak economic fluctuation. This is another indication that our model choice is correct.

Keywords: small open economy real business cycle model, Slovak business cycles, CES production function, calibration.

JEL Classification: C44, C61, E32
AMS Classification: 62P20, 90C15

1 Introduction

In our latest research we found that in spite of the Slovak Economy is emerging, its observed business cycle characteristics better satisfy to the small open developed economy: consumption is less volatile than GDP, trade balance and current account are countercyclical [12]. Furthermore we rejected the impact of policy inconsistency and interest-rate shocks on the short-run and medium-run Slovak economic behaviour [15], while Uribe and Grohé-Schmitt [16] state that both these elements can in generally significantly explain the business cycle of the emerging countries.

Therefore we use Mendoza's [6] small open economy real business cycle model modified by Schmitt-Grohé and Uribe [8] and by Uribe and Schmitt-Grohé [16] to explain the Slovak business cycle. While the model with Cobb-Douglas production function with labour share estimated in our paper [14] does not sufficiently fit the computed second moments of output, consumption, investment, trade balance and current account cyclical components we remove the Cobb-Douglas production function by the constant elasticity of the input substitution (CES) production function in the model.

2 The Small-Open-Economy Real-Business-Cycle Model

The model presented in this section is based, to a large extent, on Schmitt-Grohé and Uribe [8],Uribe and Schmitt-Grohé [16] and Mendoza [6]. Our contribution is in removing the Cobb-Douglas production function by the CES production function. This modification improves the model predictions for the Slovak observations.

Consider a small open economy populated by an infinite number of identical households with preferences described by the GHH [2] utility function:

$$U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{G(c_t, h_t)^{1-\sigma} - 1}{1-\sigma}, \ \sigma > 0 \tag{1}$$

where c_t denotes consumption, h_t denotes hours worked, $\beta \in (0, 1)$ is the subjective discount factor. The symbol E_0 denotes the expectations operator conditional on information available in initial period 0 and

[1] University of Economics in Bratislava, Dolnozemská cesta 1, Bratislava, Slovakia, szomolan@euba.sk.

$$G(c,h) = c - \frac{h^{\omega}}{\omega},\ \omega > 1$$

where ω expresses a wage elasticity.

The period-by-period budget constraint of the representative household is given by

$$d_t = (1 + r_{t-1}) d_{t-1} - y_t + c_t + i_t + \frac{\phi}{2}(k_{t+1} - k_t)^2;\ \forall t \in \{0,1,\ldots\} \tag{2}$$

where d_t denotes the household's debt position at the end of period t, r_t denotes the interest rate at which domestic residents can borrow in period t, y_t denotes domestic output, i_t denotes gross investment, and k_t denotes physical capital. The term $\phi/2(k_{t+1} - k_t)^2$ with $\phi > 0$ is meant to capture capital adjustment costs. Small open economy models typically include capital adjustment costs to avoid excessive investment volatility in response to variations in the productivity of domestic capital or in the foreign interest rate. The restrictions imposed ensure that in the steady state adjustment costs are nil and the relative price of capital goods in terms of consumption goods is unity.

The form (2) corresponds to the centralised conditions where we assume that households owning the firms have the income of output y. It follows from (2), that trade balance defined as difference between output and domestic demand is:

$$x_t = y_t - c_t - i_t - \frac{\phi}{2}(k_{t+1} - k_t)^2;\ \forall t \in \{0,1,\ldots\} \tag{3}$$

where x_t denotes the trade balance in period t. The current account is given by the sum of the trade balance and net investment income, that is,

$$b_t = -r_{t-1} d_{t-1} + x_t;\ \forall t \in \{0,1,\ldots\} \tag{4}$$

Alternatively, one could construct the process of the current account by using the fact that the current account measures the change in net foreign assets, that is,

$$b_t = d_{t-1} - d_t;\ \forall t \in \{0,1,\ldots\} \tag{5}$$

Output is produced by means of CES production function that takes capital and labour services as inputs

$$y_t = A_t \left(\lambda k_t^{\gamma} + \mu h_t^{\gamma}\right)^{\frac{1}{\gamma}};\ \forall t \in \{0,1,\ldots\} \tag{6}$$

where

$$\lambda = \alpha \kappa^{\gamma}$$
$$\mu = \eta \nu^{\gamma}$$

Parameters α and η denote capital and labour share respectively, parameter γ expresses the constant inputs elasticity of substitution and κ and ν are normalisation parameters, A_t is an exogenous and stochastic productivity shock which represents the single source of aggregate fluctuations in the present model.

$$\ln A_{t+1} = \rho \ln A_t + \zeta \varepsilon_{t+1};\ \forall t \in \{0,1,\ldots\} \tag{7}$$

where the parameter $\rho \in (-1, 1)$ measures the serial correlation of the technology shock, the innovation ε_t is i.i.d. with mean zero and unit standard deviation, and the parameter ζ governs the standard deviation of the innovation to technology. In the small open economy model we can also interpret productivity shocks as the trade shocks which have similar effect on the economy as productivity shocks.

The stock of capital evolves according to

$$k_{t+1} = (1 - \delta) k_t + i_t;\ \forall t \in \{0,1,\ldots\} \tag{8}$$

where $\delta \in (0, 1)$ denotes the rate of depreciation of physical capital.

Households are assumed to be subject to the following sequence of borrowing constraints that prevents them, at least in expectations, from engaging in Ponzi games:

$$\lim_{j\to\infty} E_t \frac{d_{t+j}}{\prod_{s=0}^{j}(1+r_s)} \le 0 \tag{9}$$

This limit condition states that the household's debt position must be expected to grow at a rate lower than the interest rate r in the long run.

Uribe and Schmitt-Grohé [16] showed that the equilibrium of a small open economy with one internationally traded bond and a constant interest rate satisfying $\beta(1 + r) = 1$ features a random walk in consumption, net external debt, and the trade balance. Under perfect foresight, that model predicts that the steady state levels of debt and consumption depend on initial conditions, such as the initial level of debt itself. The nonstationarity of the model complicates the task of approximating equilibrium dynamics, because available approximation techniques require stationarity of the state variables. Following Schmitt-Grohé and Uribe [8] we induce stationarity by making the interest rate debt elastic. Uribe and Schmitt-Grohé [16] study various alternative ways to accomplish this task.

We assume that the interest rate faced by domestic agents, r_t, is increasing in the country's cross-sectional average level of debt, which is assumed to equal to the individual level of debt d_t. Specifically, r_t is given by

$$r_t = r^* + \psi\left(e^{d_t-\bar{d}} - 1\right);\ \forall t \in \{0,1,\ldots\} \tag{10}$$

where $\psi > 0$ and $\bar{d}$ are parameters, r^* denotes the world interest rate and second term of (10) is a country-specific interest rate premium. According to this expression the country premium is an increasing and convex function of net external debt. For simplicity, we assume that the world interest rate, r^*, is constant. The assumption of a debt-elastic interest rate premium gives rise to a steady state of the model that is independent of initial conditions. In particular, under this formulation the deterministic steady state of the model is independent of the initial net foreign asset position of the economy. This feature of the model is motivated on purely technical grounds. However, the assumption of a debt elastic interest-rate premium may be of interest also because it can capture the presence of financial frictions.

A competitive equilibrium is a set of processes $\{d_t,\ k_{t+1},\ c_t,\ h_t, A_{t+1}\}_{t=0}^{\infty}$ so as to maximize the utility function (1) subject to (2)-(9), given A_0, d_{-1}, and k_0, and the process $\{\varepsilon_t\}_{t=0}^{\infty}$. Given the equilibrium processes of consumption, hours, capital, and debt, output is obtained from production function (6), investment from equation (8), the interest rate from equation (10)**,** trade balance from equation (3) and current account from equation (4).

In the deterministic economy environment we assume that the variance of the innovation to the productivity shock, ζ, is nil. We define a deterministic steady state as equilibrium of the deterministic economy in which all endogenous variables are constant over time.

This characterization is of interest for two reasons. First, the steady state facilitates the calibration of the model. This is because, to a first approximation, the deterministic steady state coincides with the average position of the stochastic economy. In turn, structural parameters are often calibrated to target average characteristics of the economy such as labour shares, consumption shares, and trade-balance-to-output ratios. Second, the deterministic steady state is often used as a convenient point of reference around which to approximate the equilibrium dynamics of the stochastic economy.

The model is used for calibration of parameters describing cyclical properties of Slovak aggregates.

3 Methodology

We assigned values to the model's structural parameters so that the computed quantitative predictions of a business-cycle model fit with observed Slovak cyclical properties.

There are several ways to accomplish it. We use the simple one – calibration. In general, the calibration method assigns values to the parameters of the model in three different ways:

1. Based on sources unrelated to the macro data the model aims to explain.
2. To match first moments of the data that the model aims to explain.
3. To match second moments of the data the model aims to explain.

We estimated these second moments in [12]. In the present model, there are 12 parameters that need to be calibrated: σ, δ, r^*, λ, μ, γ, $\bar{d}$, ω, ϕ, ψ, ρ and ζ. We separate these parameters into these three calibration categories described above.

The parameters that fall in the first category are δ, r^*, γ, a ω. We estimated the capital depreciation rate δ in [11], its value is 6.63%. Mendoza sets r^* equal to 4 percent per year. We estimated the parameter γ that captures the input substitution elasticity in [14]. Furthermore it follows from this paper that there is only one co-integrated relation between average labour product and the real wage rate which we used for the estimate of production function parameters. We therefore assume that the effect of the real wage rate change on the labour supply is only short-term. We can support this assumption by the fact that this relationship follows from the short-run business cycle theory. Using data from our paper [14] and the least square method we estimated short-run relation in which the labour first differences are explained by the wage rate first differences. We computed the parameter ω from the estimate, ω = 28.5 in Slovak economy.

In the second category is the parameter $\bar{d}$, pertaining to the country interest-rate premium. It is set to match the observed average trade-balance-to-output ratio in Slovakia, x/y = –0.016. In the deterministic steady state all endogenous variables are constant over time therefore capital adjustment costs vanish. Setting constant capital and debt and combining (2) and (3) implies that in the steady state

$$x = r^* \bar{d}$$

We denote steady-state value of variables by removing the time subscript. This condition states that in the deterministic steady state the country must generate a trade surplus sufficiently large to service its external debt. Solving for $\bar{d}$, yields

$$\bar{d} = \frac{\frac{x}{y}}{r^*} y$$

where we divided and multiplied the right-hand side by y to express the trade balance as a share of output. At this point we know x/y and r^*, but y remains unknown. From the derivation of the steady state one can deduce y. The unknown parameters in the expression for y, and therefore $\bar{d}$, are λ and μ. Next, we discuss how the calibration strategy assigns values to them and the remaining unknown structural parameters.

The calibration strategy for the rest parameters is to match the second moments of the Slovak data at business-cycle frequency mentioned in the Table 2. These observed summary statistics are taken from our paper [12]. The table shows standard deviations (σ_{zt}), serial correlations ($\rho_{zt,zt-1}$), and contemporaneous correlations with output ($\rho_{zt,yt}$) of output, consumption, investment, hours, the trade-balance-to-output and current-account-to-output ratios. The data captured from the World Bank (WB) and SLOVSTAT (S) portals is annual, detrended by Hodrick Prescott filter, and covers the period 1992 – 2010 (World Bank) and 1997 – 2012 (SLOVSTAT) respectively. In practice, this last step of the calibration procedure goes as follows:

1. Guess values for the parameters in the third category. This automatically determines a value for $\bar{d}$.
2. Approximate the equilibrium dynamics of the model.
3. Calculate the implied second moments to be matched.
4. If the match between actual and predicted second moments is judged satisfactory, the procedure has converged. If not, try a new guess for the third category parameters to be calibrated and return to 1.

In order of finding model equilibrium in the second step we apply the first order linear approximation to the nonlinear solution. We use algorithms created and modified by Klein [5] and Schmitt-Grohé and Uribe [9]. Uribe and Schmitt-Grohé [16] in fourth chapter explained the linear approximation principles.

4 Results

The parameter values that result from this calibration procedure are shown in the Table 1.

parameter	σ	δ	r^*	λ	μ	γ
value	1.5	0.066 3	0.04	1.7	0.1	–0.426

parameter	$\bar{d}$	ω	ϕ	ψ	ρ	ζ
value	–3.747 6	28.5	0.001 3	0.000 742	0.42	0.027

Table 1 Calibrated Values of Structural Parameters

Last columns of the Table 2 also display second moments predicted by the model.

variable	σ_{zt}		$\rho_{zt,zt\text{-}1}$		$\rho_{zt,yt}$		σ_{zt}	$\rho_{zt,zt\text{-}1}$	$\rho_{zt,yt}$
z	WB	S	WB	S	WB	S		model	
y	3.55	3.81	0.46	0.62	1.00	1.00	3.73	0.67	1.00
c	2.43	3.24	0.31	0.68	0.71	0.84	2.45	0.98	0.65
i	13.97	11.88	0.36	0.23	0.66	0.78	11.98	0.06	0.70
x/y	3.04	2.04	0.22	0.06	–0.20	–0.20	3.76	0.04	–0.26
b/y		1.76		0.17		–0.36	3.693	0.02	–0.27

Table 2 Empirical and the Theoretical Second Moments

It follows that model predicts output more volatile than consumption and less volatile than investment. Consumption and investment are pro-cyclical and trade-balance-to-output ratio is weakly countercyclical. These results match the observed values. However, trade balance and current account volatilities are slightly overestimated. Theoretical serial correlation of consumption is too high, while theoretical serial correlation of investment is too low.

Figure 1 displays the impulse response functions of a number of variables of interest to a technology shock of size 1 percent in period 0. In response to this innovation, the model predicts an expansion in output, consumption, investment, and hours and deterioration in the trade-balance-to-output ratio. The level of the trade balance, not shown, also falls on impact. This means that the initial increase in domestic absorption (i.e., the increase in $c_0 + i_0$) is larger than the increase in output.

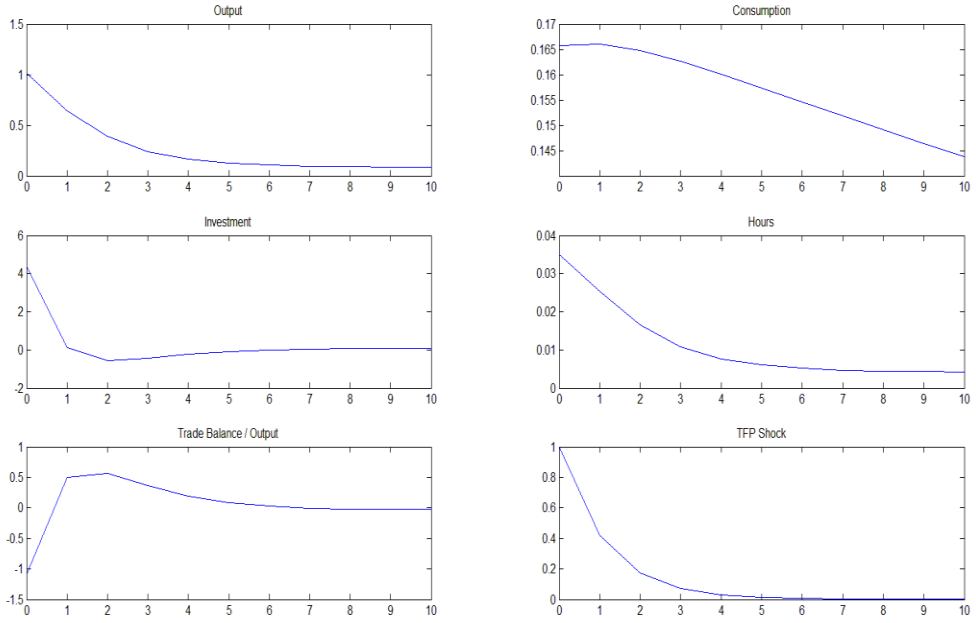


Figure 1 Responses to a One-Percent Productivity Shock

5 Conclusion

Slovak economy is small open and emerging. From results in our paper [12] we consider the model described in the second part of this paper for the best in explaining Slovak business cycle. Model provide satisfactory results if we consider the CES production function (6). In the view of the short-run economic motion, the only Slovak feature that is characteristic for the emerging countries is relatively high economic volatility and long expansion duration. Government purchases are little volatile and countercyclical, consumption is less volatile than output. Uribe and Schmitt-Grohé [16] consider interest rate shocks for significant business cycle source in Latin America. We rejected the significance of interest rate shocks in explaining Slovak business cycles in term 2001 – 2011 [15]. The time consistency of policy theory states that domestic fiscal policy has impact on the country interest rate spread. We reject this theory in the Slovak economy conditions as well.

We guess that significant source of Slovak business cycle are trade shocks. The terms of trade changes have similar business cycle impact as real shocks we considered in this paper. Great part of total factor productivity shocks in the small open real business cycle model can be interpreted as trade shocks. The unsatisfactory model explanation of the serial correlation of variables is further argument for trade shocks. It is well known that trade shocks and productivity shocks have different duration. Possible model extension differing trade and productivity shocks could improve the model results and its variable persistence predictions. Mendoza [7] introduced similar model, however Uribe and Schmitt-Grohé [16] (chapter 7) questioned part of his methodology, therefore we did not use Mendoza's model.

It is worth to note that interpretation possibilities of dynamic general equilibrium models using the economy equilibrium approximation along the steady state are low for economies that are far from the steady state point. It follows from our research [13] that Slovak economy is in transition state. This is another possible cause of the theoretical model variable persistence failure.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic - VEGA, grant no. 1/0595/11 "Analysis of Business Cycles in the Economies of the Euro Area (with Regards to the Specifics of the Slovak Economy) Using Econometric and Optimization Methods"

References

[1] Chocholatá, M.: The global financial crisis and stock returns: evidence from the Czech Republic, Hungary and Poland. In: *Mathematical methods in economics 2013: proceedings of the 31st international conference, part 2*. College of Polytechnics, Jihlava, 2013, 315-320.

[2] Greenwood. J., Hercowitz. Z. and Huffman. G.: Invesement, Capacity Utilization and the Real Business Cycle. *American Economic Review* **78** (1988). 1088-1097.

[3] Hančlová J.: Comparative Properties of Structural Macroeconomic Modelling. In: *Proceedings of the 14th International Scientific Conference Quantitative Methods in Economics: Multiple Criteria Decision Making*. High Tatras, 2008, 70-75.

[4] Hančlová J.: Panel Modelling of Globalization on Government Expenditures for the Selected New EU Countries. In: *Proceedings of the 10th International Conference of Liberec Economic Forum*, Liberec, 2011, 161-170.

[5] Klein. P.: Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model. In: *Journal of Economic Dynamics and Control* **24** (2000). 1405-1423.

[6] Mendoza. E.: Real Business cycles in a small-open economy. *American Economic Review* **81** (1991). 797-818.

[7] Mendoza. E.: The Terms of Trade, the Real Exchange Rate, and Economic Fluctuations. *International Economic Review* **36** (1995). 101-137.

[8] Schmitt-Grohé. S. and Uribe, M.: Closing Small Open Economy Models. *Journal of International Economics* **61** (2003). 163-185.

[9] Schmitt-Grohé. S. and Uribe, M.: Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function. *Journal of Economic Dyanmics and Control* **28** (2004). 755-75.

[10] Surmanová, K., Furková, A.: Methods for Output Gap Detection. In: *Mathematical methods in Economics 2011: proceedings of the 29th international conference, part 2*. Janská Dolina, 2011, 665-670.

[11] Szomolányi. K.: *Analýza slovenskej ekonomiky – teoretické východiská v rámci dynamického modelu všeobecnej rovnováhy.* Inaugural Dissertation, University of Economics in Bratislava, Faculty of Economic Informatics. 2013

[12] Szomolányi. K.: Hospodárske cykly v slovenskej ekonomike. In: *Proceedings of the AIESA*. Bratislava, 2013.

[13] Szomolányi. K., Lukáčik. M. and Lukáčiková. A.: The β-convergence of the EU27 countries. In: *Mathematical methods in Economics 2011: proceedings of the 29th international conference, part 2*. Janská Dolina. 2011, 287-291.

[14] Szomolányi. K., Lukáčiková. A. and Lukáčik. M.: The Production Function of the Transitive Economy - Case of Slovakia. In: *Book of abstracts: 26th European conference on operational research*. Roma. 2013, 385.

[15] Szomolányi. K., Lukáčiková. A. and Lukáčik. M.: Obmedzenia na svetových finančných trhoch a slovenská ekonomika. In: *Proceedings of the Inproforum 2013*. České Budějovice, 2013, 295-299

[16] Uribe, M. and Schmitt-Grohé. S.: *Open Economy Macroeconomics.* [Online] http://www.columbia.edu/~mu2166 (5th. May. 2014).

Investment decision making using fuzzy scorecards - mutual funds selection

Tomáš Talášek [1], Eva Bohanesová [2], Jan Stoklasa [3], Jana Talašová [4]

Abstract. The paper presents a multiple criteria decision support method for investment decision making. Investment decision support requires tools able to deal with imprecision, understandable to the decision makers and able to provide easily interpretable outputs. We therefore propose here a method utilizing fuzzy scorecards as an easy way to obtain expert evaluations of each alternative (mutual fund) against each criterion. Fuzzy OWA aggregation with fuzzy weights is then applied to obtain overall evaluations of alternatives in the form of piecewise linear fuzzy numbers. The final evaluations are also assigned appropriate linguistic descriptions to facilitate the decision making as well as possible. It is also possible to order the alternatives based on the centres of gravity of the overall evaluations. The mathematical model has been implemented in the FuzzME software for fuzzy multiple criteria evaluation. A practical example of mutual funds selection using this software is presented. The proposed methodology provides means of dealing with the uncertainty inherent in investment decision making and also a linguistic level of decision support for this area.

Keywords: Investment decision making, multiple criteria decision making, fuzzy scorecards.

JEL classification: 91B06
AMS classification: C44

1 Introduction

Free funds investment is still an actual topic not only among experts, but it can pertain everyone with regular incomes. Low interest rates of bank deposits and the desire for higher returns, ensuring the funds for the lives of our children or the old age, are some of the reasons. Nowadays, there are several possibilities how to increase the value of finance funds. Mutual funds represent one of them. They are usually classified depending on the kind of underlying assets or on investors risk attitude. It means that low risk funds as well as very high risk funds are offered and/or demanded. Investor making the decision about the suitable fund has to consider his/her investment aim. It means that the future need for money has to be estimated. Next, it is necessary to establish the moment in the future, when the money is to be spent. The future amount of funds is related to the profit from the investment and to the risk that the investor has to endure. The higher need for money is the higher profit can be demanded, but the higher risk has to be undergone. The moment of future funds consumption is related to investment period. Here it holds that the longer the period is the more risky investment strategy can be chosen. Investor's knowledge and experience in trading on financial markets play in the selection of the mutual fund the important role too.

The process of the mutual fund selection consists of several steps: First an investor has to know his/her investment aim and the length of the investment horizon. Second the investor has to know

[1]Palacký University in Olomouc/Faculty of Science, Department of Mathemetical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46, Olomouc, Czech Republic, tomas.talasek@upol.cz
[2]Palacký University in Olomouc/Faculty of Science, Department of Mathemetical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46, Olomouc, Czech Republic, eva.bohanesova@upol.cz
[3]Palacký University in Olomouc/Faculty of Science, Department of Mathemetical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46, Olomouc, Czech Republic, jan.stoklasa@upol.cz
[4]Palacký University in Olomouc/Faculty of Science, Department of Mathemetical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46, Olomouc, Czech Republic, jana.talasova@upol.cz

his/her risk profile, i.e. how large lost of the value of the investment is able to endure. If the investor does not know his/her risk profile, it is possible to find it out by filling in the investment questionnaire. The corresponding investment strategy is recommended to him/her on the base of evaluation of his/her answers. Now the selection of the mutual fund on the base of recommended investment strategy can start. The investor proceeds using several criteria, that usually are historical rate of return, historical volatility, Sharpe ratio, fees, possibility of their sequential repayment, economic sector diversification, geographical diversification, hedging of currency risk, reputation of the investment company etc. The criteria: *reputation of the investment company*, *amount of the fees*, *Sharpe ratio* and *currency risk* will be used in the numerical example at the end of our contribution.

The structure of the paper is following: basic notions of the theory of fuzzy sets are defined in Section 2. The use of fuzzy scorecards in decision making is briefly summarized in Section 3. In Section 4 we propose a new method for investment decision making using fuzzy scorecards to obtain expert evaluations of the variants against criteria. A practical example of a mutual funds selection problem solved using the proposed method is presented in the Section 5 and the features of the proposed method are summarized here.

2 Preliminaries

Let U be a nonempty set (the universe of discourse). A *fuzzy set* A on U is defined by the mapping $A : U \to [0, 1]$. For each $x \in U$ the value $A(x)$ is called a *membership degree* of the element x in the fuzzy set A and $A(.)$ is called a *membership function* of the fuzzy set A. $\mathrm{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an α-*cut* of A for any $\alpha \in [0, 1]$, $\mathrm{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\mathrm{Ker}(A) \neq \emptyset$; (2) A_α are closed intervals for all $\alpha \in (0, 1]$; (3) $\mathrm{Supp}(A)$ is bounded. A fuzzy number A is said to be defined on [a,b], if $\mathrm{Supp}(A)$ is a subset of an interval $[a, b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \mathrm{Cl}(\mathrm{Supp}(A))$ and $[a_2, a_3] = \mathrm{Ker}(A)$, where $\mathrm{Cl}(\mathrm{Supp}(A))$ denotes a closure of $\mathrm{Supp}(A)$. The fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$. If the membership function of a fuzzy number A is piece-wise linear on $[a_1, a_2]$ and $[a_3, a_4]$, the fuzzy number is called *piece-wise linear*. A linear fuzzy number C is said to be *triangular* if its two middle significant values are equal; for such fuzzy numbers we will use a simplified notation $C = (c_1, c_2, c_3)$, where $[c_1, c_3] = \mathrm{Cl}(\mathrm{Supp}(C))$ a c_2 is the sole element of $\mathrm{Ker}(C)$.

The following special structures of fuzzy numbers can be effectively used in the field of fuzzy multiple criteria evaluation methods: fuzzy scales on $[a, b]$ and normalized fuzzy weights. A *fuzzy scale* on $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on [a,b], that form a Ruspini fuzzy partition (see [3]) of the interval $[a, b]$, i.e. for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. Fuzzy numbers $V_1, V_2, \ldots, V_m$ defined on $[0, 1]$ form *normalized fuzzy weights*, if for any $\alpha \in (0, 1]$ and for any $i = \{1, \ldots, m\}$ the following holds: for any $v_i \in V_{i\alpha}$ there exist $v_j \in V_{j\alpha}$, $j = 1, \ldots, m$, $j \neq i$, such that $v_i + \sum_{j=1, j\neq i}^{m} v_j = 1$.

In applications an approximation of a fuzzy number A by its center of gravity t_A is frequently used. The *center of gravity* of a fuzzy number A, except for fuzzy numbers where $a_1 = a_2 = a_3 = a_4$ (so called *fuzzy singletons*), is defined by the formula $t_A = \int_a^b A(x)xdx / \int_a^b A(x)dx$. The center of gravity of a fuzzy singleton is defined as $t_A = a_1$.

3 Fuzzy scorecards and their use in decision support

Luukka and Collan in their paper [2] use fuzzy scorecards for expert evaluations of alternatives. In their approach the fuzzy evaluations take the form of triangular fuzzy numbers; the significant values of these fuzzy numbers are integers from the discrete evaluation scale $1, 2, \ldots, 10$. It is apparent from the context, that one of the extreme values of the scale represents a completely unsatisfactory evaluation and the other extreme value an entirely satisfactory evaluation. Each variant is thus expertly evaluated against all the evaluation criteria. The authors propose to standardize the evaluations of all the variants against a given criterion. This is done in such a way that all the evaluations are divided by a constant, that is equal to the maximal significant value of all the evaluations against this criterion.

The partial evaluations are aggregated into a final evaluation by a fuzzified HOWA operator [5]. In the described method only the partial evaluations are fuzzy, the weights are real numbers. Similarly to the OWA aggregation operator [6], in the HOWA aggregation the j-th weight expresses the importance of the j-th highest evaluation of the given variant. These two operators however differ in the interpretation of the weights. In the case of the HOWA operator the j-th weight expresses a necessity measure of the j-th best evaluation, while in the case of the OWA operator the weights express the shares of the respective partial evaluations of the overall evaluation. The aggregated fuzzy evaluations computed this way are then again standardized as described above. The variants are finally ordered based on the similarity of their final evaluations with an expertly defined evaluation of an ideal variant.

4 Fuzzy scorecards in investment decision making using FuzzME

Fuzzy scorecards used by Luukka and Collan can be considered as an appropriate tool for obtaining expert evaluation in the investment decision making area as well. We however propose a different approach to processing of such evaluations - an approach based on the FuzzME software [7]. FuzzME[1] constitutes a considerably universal software tool for multiple criteria decision making and evaluation. Its demo version is available at the website www.fuzzme.net. A basic feature of the evaluations with which the FuzzME-based models work is that the evaluations represent a fuzzy degree of fulfillment of an evaluation goal. This means that the evaluations are of the same type as expert evaluations obtained by fuzzy scorecards in [2]. (Let us remark here that FuzzME linearly transforms all the input evaluations to the interval $[0\,,1]$; this is however just a formal modification.)

Expert evaluations of investment variants against various criteria need not be standardized in any way, since the same evaluation scale is used for all criteria. If we consider the expert fuzzy evaluations obtained by the fuzzy scorecards, then an evaluation scale represented by a closed interval $[1\,,10]$ is used; this interval is linearly transformed to $[0,1]$ in the FuzzME representation. Since the weights assigned by experts to the evaluations are uncertain in their nature, we use a fuzzification of a higher degree (not only expertly defined evaluations but also their weights are allowed to be fuzzy). When defining the fuzzy weights expertly it is necessary to keep the structure of normalized fuzzy weights. Normalized fuzzy weights express uncertain shares of partial evaluations on the final evaluation. The FuzzME software does not include the HOWA operator in its basic selection of fuzzified aggregation operators. Nevertheless if we use normalized fuzzy weights and fuzzy evaluations of variants determined through the fuzzy scorecards and apply the fuzzified OWA operator, we obtain analogical results to those obtained by the fuzzified HOWA operator. If the fuzzy evaluations of a given variant according to a set of m criteria are described by the fuzzy numbers $U_1, ..., U_m$, then the membership function of the final fuzzy evaluation U is defined in the following way.

Definition 1. Let $U_j,\ j = 1, \ldots, m$ be fuzzy numbers on $[0, 1]$ and $V_j,\ j = 1, \ldots, m$ be normalized fuzzy weights. Fuzzy ordered weighted average of $U_1, U_2, \ldots, U_m$ with fuzzy weights $V_1, V_2, \ldots, V_m$ is a fuzzy number U with a membership function given for any $u \in [0, 1]$ by

$$\begin{aligned} U(u) \quad = \quad & \max\Big\{ \min\{U_1(u_1), \ldots, U_m(u_m), V_1(v_1), \ldots, V_m(v_m)\} \big| \\ & u_j \in [0,1],\ j = 1, \ldots, m,\ v_j \geq 0,\ j = 1, \ldots, m,\ \sum_{j=1}^{m} v_j = 1,\ u = \sum_{j=1}^{m} v_j u_{i_j} \Big\} \end{aligned}$$

where $i_1, \ldots, i_m$ denotes a permutation of indices $1, \ldots, m$ such that $u_{i_1} \geq \cdots \geq u_{i_m}$.

An effective algorithm described in [4] is used for practical computation of the fuzzy number U in FuzzME. Even in the case when all the input fuzzy numbers (normalized fuzzy weights and partial fuzzy evaluations) are linear, the final fuzzy evaluation can still be a piece-wise linear fuzzy number. The fuzzy evaluations computed this way can be interpreted in the same way as the input partial evaluations - as fuzzy degrees of fulfillment of the overall goal. The fuzzy evaluations computed by FuzzME can be easily $(y = 9\,x + 1)$ transformed back to the original evaluation scale [1,10] used in the fuzzy scorecards.

[1] developed by our research team at Palacký University, Olomouc; more detail can be found at www.fuzzymcdm.upol.cz

A significant advantage of the FuzzME software is its ability to approximate the final evaluations (piece-wise linear fuzzy numbers) linguistically. For decision makers in the area of investments a linguistic approximation by elements of a linguistic evaluation scale (whose mathematical meanings form a fuzzy scale) seems to be appropriate. It is also possible to use more complex linguistic structures derived from the scale for the linguistic approximation (see the numerical example). To compare the final evaluations, the center of gravity method can be also used.

5 Numerical example – mutual funds selection

The investment decision support method proposed in the previous section will be illustrated here on a case of recommending the best of four mutual funds. Four suitable mutual funds are considered: Pioneer Akciový fond *(Pioneer)*, ISČS Top stocks *(ISČS)*, ČPI fond globálních značek*(ČPI)* and C-QUADRAT Strategie AMI CZK *(C-QUADRAT)*. Four criteria are considered in the evaluation of each fund: Reputation of the investment company (C_1), Sharpe ratio (C_2), Amount of fees (C_3), Currency risk (C_4).

The reputation of the investment company conducting collective investment is important. None would invest money to an incredible company. Thus the investor will prefer company with the best reputation. Investments to mutual funds are not usually without fees. These represent commissions for agents and remunerations for portfolio managers. Their amount depends on the investment strategy: the more risky the strategy is the higher fees have to be paid. Naturally, the investor prefers lower fees to higher ones, but if he/she wants to reach higher returns and invests to more risky funds, it is worth higher fees. Sharpe ratio is the indicator that expresses the rate of return after benchmark subtraction to the unit of the risk (measured by standard deviation). Higher values of the ratio are preferred. If funds are invested to foreign securities, currency risk is expected. Portfolio managers prevent from the risk by hedging. Mutual funds with hedged currency risk are recommended to investors.

The preference system of the decision maker is reflected in the assignment of weights in the fuzzy OWA method (higher weights are assigned to worse fuzzy evaluations of the fund, a conservative decision maker is considered; the actual weighting vector applied in this example is $w = [(0.1, 0.2, 0.3), (0.13, 0.23, 0.33), (0.17, 0.27, 0.37), (0.2, 0.3, 0.4)]$). Table 1 summarizes the four mutual funds under consideration and the four characteristics based on which the expert evaluation is carried out (TER - total expense ratio in % p.a., CRH - currency risk hedge). The evaluation is performed using fuzzy scorecards. Table 2 summarizes the evaluations of the mutual funds under consideration obtained by fuzzy scorecards.

Mutual fund	Investment company	Sharpe ratio	TER	CRH
Pioneer	Pioneer investiční společnost, a.s	0.61	2.29%	partially
ISČS	ČP Invest investiční společnost, a.s.	0.81	2.49%	yes
ČPI	Investiční společnost ČS, a.s.	0.91	2.79%	yes
C-QUADRAT	C-QUADRAT Kapitalanlage AG	0.76	3.76%	yes

Table 1 Characteristics of mutual funds in the numerical example.

Mutual fund	C_1	C_2	C_3	C_4
Pioneer	(3, 5, 7)	(1, 2, 4)	(6, 8, 9)	(2, 3, 5)
ISČS	(8, 9, 10)	(6, 8, 9)	(4, 5, 7)	(9, 10, 10)
ČPI	(6, 8, 9)	(5, 7, 8)	(5, 7, 8)	(9, 10, 10)
C-QUADRAT	(5, 7, 8)	(2, 3, 6)	(2, 3, 5)	(9, 10, 10)

Table 2 Funds ratings - summary of evaluations for all criteria.

A screenshot from the FuzzME software (Figure 1) illustrates the normalized fuzzy weights and the fuzzy evaluation of the fund ISČS Top stocks. Let us remark that the decision maker does not need to strictly fulfill the relatively difficult-to-fulfill condition of normality of the weights. It is sufficient if the elements in the kernels of the triangular fuzzy numbers form an m-tuple of normalized real weights. The FuzzME software is capable of reducing the uncertainty of the proposed fuzzy weights so that they form

normalized fuzzy weights. In the left part of Figure 1 we can see the aggregated fuzzy evaluation, below it its center of gravity and above it its linguistic approximation.

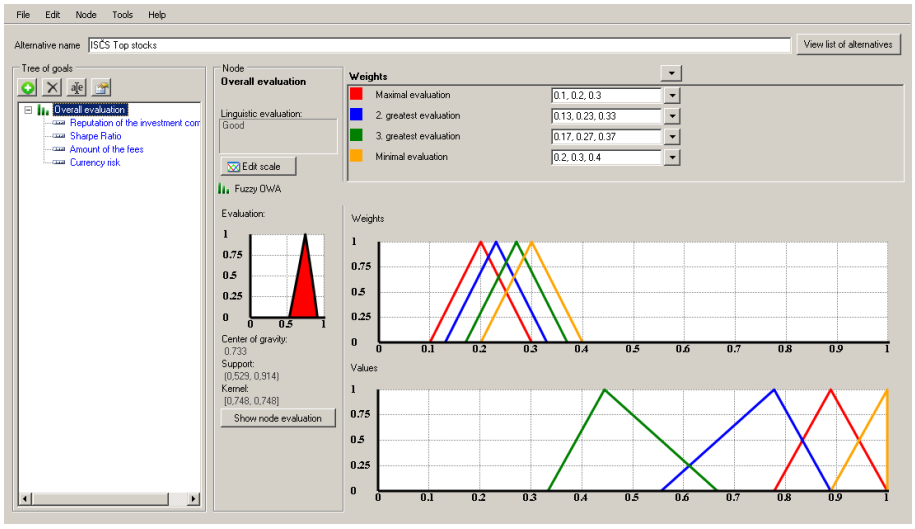


Figure 1 Evaluation overview of the ISČS Top stocks fund in FuzzME.

Figure 2 presents the linguistic evaluation scale (the meanings of its linguistic terms form a fuzzy scale) that is used as a basis for linguistic approximation of the final evaluations. It is apparent (see Table 3) that a slightly richer linguistic scale was used to obtain the linguistic approximations of the final evaluations - a so called linguistic scale with intermediate values. Final evaluations transformed back to the interval $[1, 10]$ used in the fuzzy scorecards are depicted in Figure 3. The results of the computations are summarized in Table 3.

Figure 2 Linguistic evaluation scale used in the numerical example.

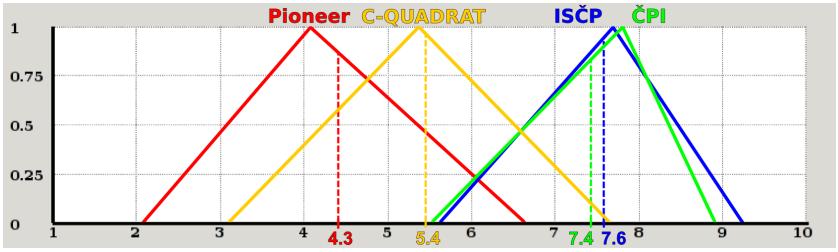


Figure 3 Visualization of final evaluations with depicted centers of gravity.

The final step - that is the actual selection of the best fund - is left to the investor. The information is therefore provided in a numerical form (centers of gravity of the resulting fuzzy evaluations), graphical outputs (membership functions of the resulting fuzzy evaluations) and linguistic outputs (linguistic approximations of the resulting fuzzy evaluations). A financial expert we have addressed interpreted the

Alternatives	Overall evaluation	Linguistic approximation	Center of gravity
Pioneer	(2.134, 4.159, 6.6610)	Between Bad and Average	4.3030
ISČS	(5.761, 7.732, 9.226)	Good	7.597
ČPI	(5.527, 7.831, 8.929)	Good	7.444
C-QUADRAT	(3.088, 5.320, 7.660)	Average	5.383

Table 3 Final evaluations of the four mutual funds.

results of this short example in the following way: ISČS and ČPI are both considered to be good choices for investment - these mutual funds (as well as their final evaluations) are fairly similar, while the Pioneer was considered to be the worst choice of the four considered mutual funds. The outputs of the model were in good accordance with the opinion of the expert. Although the inputs are obtained using as simple method as fuzzy scorecards, the obtained outputs seem to provide an appropriate basis for investment decision making (the results are well understandable). This in our opinion makes the proposed method a good candidate for a widely used decision support tool in the area of investment decision making.

6 Conclusion

A multiple criteria decision making method for investment decision making that uses fuzzy scorecards to obtain expert evaluations of the alternatives against all the criteria has been presented in this paper. The method employs a versatile multiple criteria evaluation software FuzzME, that is decision maker friendly and has all the features necessary for the decision support suitable for financial investment - mathematically correct model, where a decision maker enters input information in the way natural for him/her, and outputs, that are obtained in the understandable form. The uncertainty of information is treated in the proposed model as well as possible taking into account the needs of investment decision making.

Acknowledgements

Supported by the grant IGA_PrF_2014028 *Mathematical models* of the Internal Grant Agency of Palacký University in Olomouc and by the grant GA 14-02424S *Methods of operations research for decision support under uncertainty* of the Grant Agency of the Czech Republic.

References

[1] Collan, M.: Fuzzy or linguistic input scorecard for IPR evaluation. *Journal of Applied Operational Research*, **5**(1) (2013), 22–29.

[2] Luukka, P. and Collan, M.: Fuzzy scorecards, FHOWA, and a new fuzzy similarity based ranking method for selection of human resources. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013, 601–606.

[3] Ruspini, E.: A New Approach to Clustering. *Inform. Control*, **15** (1969), 22–32.

[4] Talašová, J. and Bebčáková, I.: Fuzzification of aggregation operators based on Choquet integral. *Aplimat - Journal of Applied Mathematics*, **1** (2008), 463–474.

[5] Yager, R. R.: Heavy OWA operators. *Fuzzy Optimization and Decision Making,* **1**(4) (2002), 379–397.

[6] Yager, R. R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics,* **18**(1) (1988), 183-190.

[7] Talašová, J., Holeček, P.: Multiple-Criteria Fuzzy Evaluation: The FuzzME Software Package. In *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference, Lisbon, Portugal*, July 20-24, 2009, 681–686.

HR management through linguistic fuzzy rule bases - a versatile and safe tool?

Jana Talašová [1], Jan Stoklasa [2], Pavel Holeček [3]

Abstract. In this paper we advocate the use of linguistic fuzzy rule-based classifiers in HR management, we discuss their possible benefits and shortcomings both from practical and from the mathematical point of view. We show that many decision tasks in HR management can be formulated in the language of classification and represented by fuzzy rule-based models. We illustrate the versatility of these models on several examples of decision making and evaluation problems. A major advantage of the linguistic rule base representation of the decision making process is its clarity and easy understandability for all people involved in the HR management process including the staff. Expert knowledge and experience can be reflected, modifications of the models are possible in real time and can be done by the practitioners. We discuss several examples of the use of linguistic fuzzy modeling in HR management in the context of interpretability and comprehensibility of the results provided by the mathematical models and possible risks and misinterpretations. The analysis was performed from the behavioral operations research point of view.

Keywords: fuzzy rule base, linguistic scale, classification, HR management, decision support, evaluation.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Human resource (HR) management involves a great deal of decision making and evaluation. Available inputs are usually numerous, their character however ranges widely from crisp numerical values to linguistic (narrative) pieces of information. As human beings are involved and the well being of the organisation in question has to be considered, context of the decision making plays an important role as well. Decision support tools able to provide quick summaries and to preprocess the vast data available can improve the HR management processes significantly. This claim can be supported by the vast literature on decision support systems (DSS) for management including HR management that have been developed recently. It is important to stress that in the area of HR management, as well as in other fields of human science, the outputs provided by the DSS and other mathematical tools need to be easy to understand and interpret. The more complex or abstract the outputs are, the more misunderstanding and misuse of these results we can expect from practitioners.

Providing preprocessed information in an easy to understand form while still retaining all the uncertainty stemming from the inputs (as in human systems there usually is uncertainty) is necessary. Mainly to enable the HR managers to concentrate on the interpretation of the preprocessed data and incorporating context and relevant soft variables into the decision process. From this perspective fuzzy rule based classification or decision support is not intended to replace the human factor in HR management. On the contrary - its proper use may enhance the potential of effective HR management by allowing the HR professionals to concentrate on issues that really require their skill, expertise and above all that

[1]Palacký University in Olomouc/Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46, Olomouc, Czech Republic, jana.talasova@upol.cz
[2]Palacký University in Olomouc/Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46, Olomouc, Czech Republic, jan.stoklasa@upol.cz
[3]Palacký University in Olomouc/Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46, Olomouc, Czech Republic, pavel.holecek@upol.cz

are directly linked with their responsibility. In our opinion any mathematical model used in the HR management context should only provide support for the decision making, not the final decisions.

All the decisions regarding people (and the well being of the company/institution) should be made by people (we need to admit that mathematical models still can not reflect the whole context of the situation adequately) and the responsibility for the consequences of their decisions should never be taken from the managers by the use of mathematical models. To achieve this, the mathematical models used to "ease the decision making process" need to provide the results carefully, with proper explanations and with a clear information on their actual (un)certainty attached. The role of the decision maker and his/her goals (as well as the goals of the organisation he/she represents) should not be neglected as well. The ethical aspects of decision support, responsibility for the decisions, comprehensibility of models for practitioners, the soundness of assumptions concerning the models - all these issues deserve appropriate attention of researchers in the field of operations research. In accordance with the ideas of the newly forming branch of OR - behavioral operations research (see [2]), we would like to address these issues here, and also discuss the advantages, as well as possible shortcomings of using advanced mathematical decision support tools such as linguistic fuzzy modeling in HR management.

2 Preliminaries

We will briefly summarize here the key concepts of fuzzy set theory and linguistic fuzzy modelling as introduced in [8] and [7] (see e.g. [1]). A (type-1) *fuzzy set* A on a universal set U can be defined by the mapping $A : U \to [0, 1]$. The symbol $\mathcal{F}(U)$ denotes the *family of all fuzzy sets* on U. $\mathrm{Ker}(A) = \{x \in U \mid A(x) = 1\}$ denotes a *kernel* of A. For any $\alpha \in [0, 1]$, $A_\alpha = \{x \in U \mid A(x) \geq \alpha\}$ denotes an *α-cut* of A. A *support of* A is defined as $\mathrm{Supp}(A) = \{x \in U \mid A(x) > 0\}$. The symbol $\mathrm{hgt}(A)$ denotes a *height of a fuzzy set* A, $\mathrm{hgt}(A) = \sup\{A(x) \mid x \in U\}$. If the *support of* A is a finite set, $\mathrm{Supp}(A) = \{x_1, \ldots, x_k\}$, then the fuzzy set A will be denoted as $A = \{{}^{A(x_1)}\!/_{x_1}, \ldots, {}^{A(x_k)}\!/_{x_k}\}$. A *union* of fuzzy sets A and B on U is a fuzzy set $A \cup B$ on U with the membership function for all $x \in U$ given by $(A \cup B)(x) = \max\{A(x), B(x)\}$. An *intersection* of fuzzy sets A and B on U is a fuzzy set $A \cap B$ on U with the membership function for all $x \in U$ given by $(A \cap B)(x) = \min\{A(x), B(x)\}$. Let A be a fuzzy set on U and B be a fuzzy set on V. Then the *Cartesian product* of A and B is the fuzzy set $A \times B$ on $U \times V$ with the membership function defined for all $(x, y) \in U \times V$ by $(A \times B)(x, y) = \min\{A(x), B(y)\}$.

A *fuzzy number* is a fuzzy set C on the set of all real numbers $\Re$ which satisfies the following conditions: a) $\mathrm{Ker}(C) \neq \varnothing$, b) C_α are closed intervals for all $\alpha \in (0, 1]$ and c) $\mathrm{Supp}(C)$ is bounded. If $\mathrm{Supp}(C) \subseteq [a, b]$ we call C a *fuzzy number on* $[a, b]$. A fuzzy set $C = \{1/_{x_1}\}$ is called a *fuzzy singleton* and represents the crisp value $x_1 \in \Re$ as a fuzzy number. A *fuzzy scale on* $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, ..., T_s$ on $[a, b]$ that form a Ruspini fuzzy partition (see [4]) of the interval, i.e. for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering $(T_1 < T_2 < \cdots < T_s)$.

A *linguistic variable* (see [7]) is defined as a quintuple $(\mathcal{X}, \mathcal{T}(\mathcal{X}), U, G, M)$, where $\mathcal{X}$ is the name of the variable, $\mathcal{T}(\mathcal{X}) = \{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_s, \}$ is a set of its linguistic values (linguistic terms), U is a universal set on which the meanings of the linguistic terms are modeled, G is a syntactic rule for generating linguistic terms from $\mathcal{T}(\mathcal{X})$, and M is a semantic rule which to every linguistic term $\mathcal{A} \in \mathcal{T}(\mathcal{X})$ assigns its representation, $A = M(\mathcal{A})$, which is a fuzzy set on U. For better clarity, from now on we will distinguish a linguistic term $\mathcal{A}$ from its mathematical meaning A, which is a fuzzy set, by a different font. In real-life applications, the universe U is usually a closed interval of real numbers, i.e. $U = [a, b]$, and the meanings of the linguistic terms are fuzzy numbers on U. A *linguistic scale* is a special case of a linguistic variable, the fuzzy numbers $T_1, T_2, ..., T_s$, representing meanings of its linguistic values, form a fuzzy scale on $[a, b]$. Let us now consider m linguistic scales $(\mathcal{E}_j, \mathcal{T}(\mathcal{E}_j), [p_j, q_j], M_j, G_j)$, $j = 1, \ldots, m$. Let $\mathcal{A}_{i,j} \in \mathcal{T}(\mathcal{E}_j)$ be linguistic values, and $A_{ij} = M_j(\mathcal{A}_{i,j})$ fuzzy numbers on $[p_j, q_j]$ representing the meanings of the respective linguistic variable's values. We can now define the relationship between the values of the linguistic variables $\mathcal{E}_j$ and an output linguistic variable $(\mathcal{H}, \mathcal{T}(\mathcal{H}), U_\mathcal{H}, M_\mathcal{H}, G_\mathcal{H})$ with values $\mathcal{D}_i \in \mathcal{T}(\mathcal{H})$ by a *fuzzy rule base* in the following form:

If $\mathcal{E}_1$ is $\mathcal{A}_{1,1}$ and ... and $\mathcal{E}_m$ is $\mathcal{A}_{1,m}$, then $\mathcal{H}$ is $\mathcal{D}_1$
If $\mathcal{E}_1$ is $\mathcal{A}_{2,1}$ and ... and $\mathcal{E}_m$ is $\mathcal{A}_{2,m}$, then $\mathcal{H}$ is $\mathcal{D}_2$
.....
If $\mathcal{E}_1$ is $\mathcal{A}_{n,1}$ and ... and $\mathcal{E}_m$ is $\mathcal{A}_{n,m}$, then $\mathcal{H}$ is $\mathcal{D}_n$.

The input linguistic variables are frequently linguistic scales. The output linguistic variable can also be a linguistic scale or a more complex structure based on a linguistic scale; the meanings of its linguistic terms can alternatively be modelled by fuzzy singletons defined on a subset of integers (these fuzzy singletons can represent the indices of linguistic terms of $\mathcal{H}$; there might exist a natural ordering of the linguistic terms, but the set of linguistic terms can be unordered as well). In HR management decision support a linguistic scale as an output variable in the fuzzy rule base corresponds with an evaluation on a cardinal scale. The case of ordered linguistic terms represented by fuzzy singletons corresponds with an evaluation on an ordinal scale and unordered linguistic terms represented by fuzzy singletons correspond with a nominal evaluation scale. Examples of such applications are provided in this paper. There are many approaches how to construct the mathematical model of each rule, the whole rule base and also how to compute and output D for an input vector $(A_1, \ldots, A_n) \in \mathcal{F}([p_1, q_1] \times \cdots \times [p_m, q_m])$. Based on the purpose and the required character of the evaluation scale a proper inference mechanism has to be chosen to compute outputs for given inputs from the fuzzy rule base. In this particular paper we aim to deal with more general issues and hence we refer the reader to the numerous books and papers on fuzzy rule bases for more details.

3 Practical applications

We stress the linguistic approach here in the context of HR management, as it provides tools for easy information transfer and for building interfaces that are easy to understand for people with no (or just a basic) mathematical background. We need to understand that decision support systems in these fields have to provide self explanatory and unambiguous results, that are not fuzzier (more uncertain) nor more precise than can be actually inferred based on the quality of the inputs (and other variables influencing the decision process, including the precision of the goals set and so on). Among the HR management tasks that can really benefit from linguistic fuzzy decision support tools, we can identify several categories, that will be briefly discussed in the following sections.

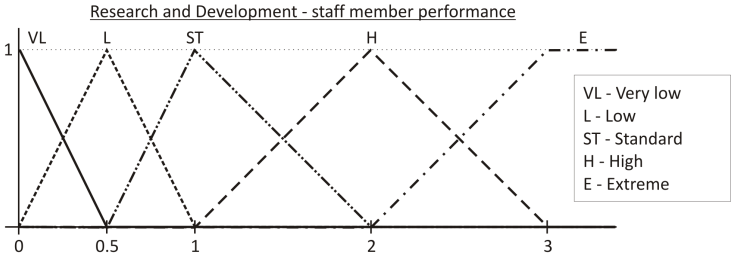


Figure 1 Meanings of the values of a linguistic scale describing performance of a worker in the area of Research and Development. Units on the x-axis are multiples of a standard performance score.

3.1 Data preprocessing

Applications of linguistic fuzzy modeling here can include transforming available data into a more manageable format and identifying relevant or key features in the data. This points to linguistic granulation, use of linguistic variables as well as to the identification of important patterns (or indicators) in data using linguistic fuzzy rules (e.g. answers to "do we need to consider promotion of this employee?", "how much are the data complete/consistent?"). Fuzzy classification can also be used in this context to point to the linguistic granule (e.g. a value of a linguistic variable describing a given aspect e.g. "very consistent") that best suits the data. Important features characterizing a given person in linguistic terms can thus be made available to the HR specialist. Linguistic granulation can also help to describe the evaluation and management process to the employees. Linguistic level of modeling provides a good communication platform between HR managers and the models, but also between HR managers and the people that are being managed. Figure 1 provides an example of a linguistic granulation of "staff member's performance in research and development". We can see that only 5 levels of performance are distinguished. Although it might be possible to compute precise numbers of the "multiples of standard score", it would be impractical to use these when providing descriptions of e.g. an evaluation procedure. By identifying 5 levels of performance, we can assign each of them a future intervention (e.g. a warning, or some bonus).

Obviously the meanings of the linguistic terms are not defined as crisp and overlap partially. This reflects the uncertainty of linguistic description (of the meaning of words). However when using linguistic labels, we need to make sure that the meanings of these labels are at least known to (and ideally also accepted and understood by) all the people using them. If the meanings of the linguistic terms are defined in a way that contradicts their common use or intuition, the changes made to a model on a linguistic level may result in an unexpected behavior of the mathematical level of the model.

3.2 Categorization

Based on the predefined (and granulated) characteristics, a classification task can be easily formulated by specifying linguistic (fuzzy) rules. Staff promotion, choosing new employees, selecting members for a new team - all these can be formulated as classification tasks (see e.g. [9] for an application from HR management in a software company). Software tools capable of implementing such rules already exist (see e.g. [3, 6]) or can be easily custom made. The use of linguistic fuzzy rule bases provides not only a high level of comprehensibility due to the linguistic level of description, but also easy adjustability of the model. Hence the tools can be easily adapted to emerging situations and to meet new requirements, goals of the organisation etc. The two tables in Figure 2 provide an example of how easy it might be to alter our model by simply changing the linguistic rules (two linguistic rule bases and two corresponding evaluation functions). The case here is to determine, to which performance class a current staff member belongs - 5 linguistically labeled classes are distinguished (from *unsatisfactory* to *excellent performance*). Using such linguistic fuzzy rule bases, a staff member can belong to more than one class (even in a different degree of membership). Also here we need to be careful and always provide meanings of the linguistic terms that are being used (see e.g. Figure 1 for the meanings of R&D performance linguistic values). We need to be aware that meanings are context dependent (the meaning of a linguistic term in a scale can even change, if we add more values to the scale and apply it to the same problem - consider a term set {young, old} and a term set {young, middle_aged, old} for the description of age; will the meanings of "young" and "old" be the same using these two term sets?). Using a developed model in a new context therefore always requires revision and analysis of its appropriateness. This requires an "informed" user or an expert to supervise changes in such models. Although linguistic representation is appealing, it is too easy to forget about context dependency of meanings. Experts designing such models should therefore be able to explain the possibilities (and restrictions) for adaptation and possible risks connected with changes of the model to the model users. Otherwise the responsibility for any mistakes is at least shared by the expert who designed the initial model. We can not provide people with tools they do not sufficiently understand and act as if the misuse of these methods is their problem and responsibility. If changes to the model might result in problematic functionality of the model, the changes should not be allowed. Let us consider Figure 2 to describe two variants of an evaluation situation (see also the next subsection), and let the outputs be computed in the following way (see [5] for more details):

$$eval(pa, rd) = \frac{\sum_{i=1}^{n} A_{i1}(pa) \cdot A_{i2}(rd) \cdot ev_i}{\sum_{i=1}^{n} A_{i1}(pa) \cdot A_{i2}(rd)} = \sum_{i=1}^{n} A_{i1}(pa) \cdot A_{i2}(rd) \cdot ev_i, \tag{1}$$

where $eval(pa, rd)$ is the output value (specifically here a real number from $[0, 2]$; pa is a multiple of standard score in the first area, rd in second), $A_{i,1}$ are the meanings of linguistic values *very low* to *extreme* describing performance in the first area of interest ($A_{i,2}$ analogically for another area of interest - see Figure 1) and the corresponding fuzzy scale and ev_i describes an output of i-th rule (crisp numbers are considered in Figure 2 as the meanings of the values of the output variable). Figure 2 then allows us to see what changes are made to the evaluation function by changing the linguistic rules. If these changes reflect well the intentions of the evaluator, everything is fine. The question is - Since we are not aware of the shape of our evaluation function when constructing the rules - does it really reflect our intentions well? And is a practitioner inexperienced with interpretation of graphs capable of assessing this? In our opinion after proper explanation the answer is yes - yet the explanation must take place and has to be done so that the model is well understood by those using it. "Black boxes" might lead to unpredictable results if modifications by the user are allowed.

3.3 Evaluation

If the linguistic values on the consequent parts of the rules allow us to asses the fulfillment of a given goal, the fuzzy rule base represents an evaluation function (see e.g. [5]). Using e.g. a Ruspini fuzzy partition to model the meanings of the linguistic terms of the output variable (see the meanings of *very low* to *extreme* in Figure 1), we obtain an evaluation scale. The linguistic terms and their meanings are ordered. Each value from the universe (described in terms of multiples of standard performance scores) can be interpreted using either single, or two neighboring linguistic terms. In fact each value is sufficiently described by its membership degrees to all the meanings of linguistic terms. A performance of a person can thus be described as e.g. 25% *standard* and 75% *high* (in fact this can also be interpreted that the performance is somewhere between *standard* and *high* and its compatibility with the label *high* is larger). This provides an easy to understand, yet uncertain, linguistic output - a valuable information for the evaluation. Alternatively as in (1) the meanings of the linguistic terms of the output variable can be represented by fuzzy singletons (with kernels ev_i) forming a uniform evaluation scale to obtain outputs of this kind.

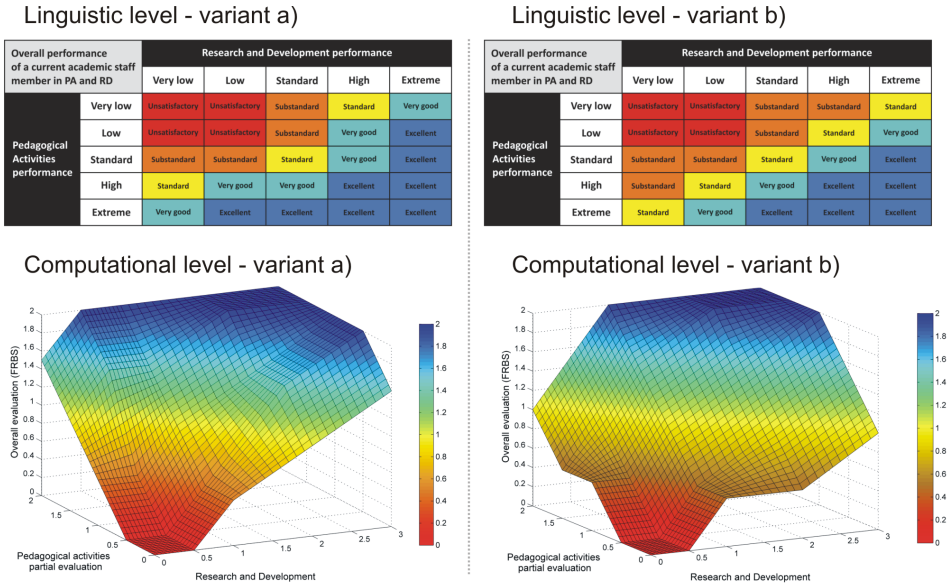


Figure 2 Example of two fuzzy rule bases differing only in the output (consequent) parts of the rules. The change of the mathematical model - the evaluation function - caused by changes in linguistic rules is obvious. Note that a complex evaluation function is described in both cases by 25 easy to understand linguistic rules and the transition from variant a) to variant b) requires no mathematical knowledge from the practitioner (user of the model). See [5] for details concerning the computation or the linguistic scales used to derive the outputs, the scale for Research and Development is depicted in Figure 1.

3.4 Data interpretation

The above mentioned outputs open a way to graphical representation of evaluation results (see e.g. [5]). In many cases we do not need precise results and more importantly precise numerical results are in fact determined based on imprecise inputs and rules. As such these should not be used or interpreted as precise. New ways of presenting outputs of fuzzy (linguistic) models might be useful to help people understand the uncertainty of the outputs, the need to interpret them in context of additional data and so on. Graphical representation using colors seems to be a promising first step.

In general using linguistic rules to describe the desired procedure, we can obtain results that are fuzzy sets with known linguistic labels assigned to them (these can be interpreted easily supposed that the labels and their meanings were defined in cooperation with the user of the model). More frequently we get fuzzy sets that partially overlap with one or more terms of our scale. In this case some procedure of assigning a linguistic label to them is necessary. Many methods were proposed in the literature, involving concentrations and dilatations of the meanings of the linguistic terms (= use of linguistic hedges such

as *very* or *somewhat)*, the use of intermediate values, finding the label whose meaning is closest to the obtained fuzzy set and so on. We need to realize that these methods result in an approximate result. Also here we need to be able to find a reasonable tradeoff between approximation and precision (although e.g. "*very [between (more or less standard) and very high*]" is a linguistic description, we need to be sure that the user is able to use it and derive information from it). Finding methods of presenting interpretable outputs without unnecessary loss of information is an important objective to be addressed now.

4 Conclusions

To answer the question posed in the title of this paper, we have presented several examples of possible uses of fuzzy rule based models in HR management setting. The usefulness and versatility of these models in HR management setting is, as we hope, apparent. The question whether these models are safe to use by laymen is, however, not so easy to answer. At least basic knowledge of the concepts of linguistic fuzzy modeling is required to be able to find problems or discrepancies between the model and reality. Fortunately the basic concepts of linguistic fuzzy modeling can be explained to laymen in reasonable time. Not doing so can, however, lead to misuses of our models and misinterpretations of their results. It is even more important when providing outputs of these models in the form of fuzzy sets. Although fuzzy outputs can carry all the information (e.g. concerning the uncertainty of the output), these are not always elementary to interpret. The price for providing easy to understand outputs is apartial loss of information (precision loss during linguistic approximation, apparent precision increase during defuzzification and so on). Graphical and linguistic outputs can be seen as a compromise between the loss of information and easy interpretability of the outputs. The tradeoff should, however, be studied more in the future and the implications of context dependency and "(the lack of) mathematical skills" of model users carefully investigated. Good explanation skills are also required when making models for practitioners, and as such should be encouraged and developed during the education of young model designers.

Acknowledgements

This research was supported by the grant *GA 14-02424S* of the Grant Agency of the Czech Republic.

References

[1] Dubois, D., and Prade, H.: *Fundamentals of fuzzy sets*. Kluwer Academic Publishers, Dordrecht, 2000.

[2] Hämäläinen, R. P., Luoma, J., and Saarinen, E.: On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems, *European Journal of Operational Research*, **228**(3) (2013), 623–634.

[3] Holeček, P., and Talašová, J.: Designing Fuzzy Models of Multiple-Criteria Evaluation in FuzzME Software, *Proceedings of the 28th International Conference on Mathematical Methods in Economics 2010*, (2010), 250–256.

[4] Ruspini, E.: A New Approach to Clustering. *Inform. Control*, **15** (1969), 22–32.

[5] Stoklasa, J., Talašová, J., and Holeček, P.: Academic staff performance evaluation variants of models, *Acta Polytechnica Hungarica*, **8** (3) (2011), 91–111.

[6] Talašová, J., and Holeček, P.: Multiple-Criteria Fuzzy Evaluation: The FuzzME Software Package, *proceedings of the Joint International Fuzzy Systems Association World Congress/European Society for Fuzzy Logic and Technology*, (2009), 681–686.

[7] Zadeh, L.A: The concept of linguistic variable and its application to approximate reasoning, *Information sciences*, **8** (1975), 199–249.

[8] Zadeh, L. A.: Fuzzy sets. *Inform. Control*, **8**(3) (1965), 338–353.

[9] Zemková, B., and J. Talašová. 2011. Fuzzy Sets in HR Management. *Acta Polytechnica Hungarica,* **8** (3): 113-124.

Modelling of Destination Portfolio for International Regional Airport

Dušan Teichmann[1], Michal Dorda[2], Vladimír Smrž[3], Jakub Vítek[4]

Abstract. Every international regional airport endeavors to operate at least one international air route to a more important airport (further transfer airport only). From the transfer airport the passengers usually travel to farther destinations (final destinations) but it is not effective for the solved international regional airport to operate the separated air routes from the regional airport to the final destinations. For solving the defined problem the linear mathematical model is proposed in the contribution. Inputs to the model include a set of transfer airports, a set of final destinations, and the demand for transport to the final destinations. The demand for transport to the final destinations is divided according to partial planning periods and is modelled using fuzzy numbers. An objective function models the total satisfied demand for transport to the final destinations on the pre-defined levels of satisfaction.

Keywords: mathematical programming, mathematical model, flights planning, regional airport**.**

JEL Classification: C61
AMS Classification: 90C90

1 Introduction – our motivation to solve the problem

For many regional international airports in Central and East Europe it is specific that their operation is irregular during a year. Most of airports' clients use the airports only during the summer season to fly to/from seaside resorts. The number of checked-in passengers is essentially lower off season than during the summer season. The management of the airport tries to compensate the loss of the passengers by offering some regular flights to selected important international airports (we will call them the transfer airports). The passengers have the possibility to fly from the transfer airports to other important airports of the world (we will call them the final destinations).

The task how to choose suitable air routes that should be operated is often solved chaotically without using any methods supporting such decisions. Therefore we have started to analyze the problem and to find out a suitable method that could make the decision easier.

The paper follows from the problem which was described in [5]. The paper contains a basic formulation of the problem and several variants of a mathematical model for its solving are presented. However, the original models have two basic lacks. The first one is that the models assume cumulated passenger flows during a certain period. The second one is that the models proposed in [5] are deterministic as regards the demand for transport. That means entry data represented by values of the demand for transport must be exactly defined and this can be a serious problem in practice. Therefore we would like to eliminate the lacks in this paper. In the paper we assume a modification of the model for which it holds as follows:

- The demand for transport to the final destinations is approximately known.
- It is presumed that at least one final destination can be reached from at least two transfer airports.
- The transfer airports also generate the demand for transport.

2 The state of the art

The state of the art can be analysed from two points of view – from the point of view of existing models to solve the problem and from the point of view how to deal with uncertainty in mathematical modelling.

[1] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 324 575, e-mail: dusan.teichmann@vsb.cz.
[2] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 325 754, e-mail: michal.dorda@vsb.cz.
[3] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 325 209, e-mail: vladimir.smrz@vsb.cz.
[4] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 321 765, e-mail: jakub.vitek@vsb.cz.

There are a lot of methods for solving the problem of designing airlines network. If we want to find a new destination for our international regional airport we can use many ways to do it. One of these ways uses a model of an analysis of a catchment area and estimates a passenger flow. The analysis of the catchment area is the analysis of the economic situation in our region, tourism, demographic data, competition of our area and a spectrum of business passengers. Then we can use for example a gravity model published in publication [7]. For solving this problem a lot of mathematical models, especially mathematical programming models, exist. One of these models was published in [6]. The author solves in his contribution how to operate a destination with different number of passengers. This problem is quite similar like our problem. The author does not forget during scheduling about different types of aircraft too. Other authors [1] solve this problem with linear programming. One of these ways is to maximize the profit on one way to the optimal destination. Other ways are: to minimize track of the aircraft, to minimize the number of the aircraft on line and to minimize the number of crew operates on one destination.

Our problem will be solved in the condition of the uncertainty. Some methods how to deal with uncertainty in mathematical modelling are described in [2], [3] or [4]. One approach is based on fuzzyfication of the existing mathematical models. This approach is quite easy in comparison with other methods. The approach does not require (in comparison with stochastic programming) more detailed information about stochastic behavior of the process we model (it is hard to make such presumptions when a new air route is formed). Only interval estimations of model's coefficients must be defined to apply fuzzy modelling. In the book [4] four basic approaches to fuzzyfication of the linear models are discussed. The first approach is used if fuzzy numbers are in an objective function. The second one arises when the fuzzy numbers are used in coefficients on the left-hand side of constraints. The third one is applied if the fuzzy numbers are used in coefficients on the right-hand side of the constraints. And finally, the last approach we apply in situations if fuzzy relation operators are used in the model. Each of the situations requests an individual method for fuzzyfication. If the fuzzy numbers are used in several parts of the model (for example both in the objective function and in the coefficients on the right-hand side of the constraints), the approaches can be combined with one another. Choosing the suitable approach to fuzzyfication will be account for in the section devoted to the mathematical model.

3 The mathematical model

Let us consider that we have an airport from which we would like to operate maximally $p \geq 2$ air routes, where $p \in Z_0^+$ and $p \in Z_0^+$ corresponds to the set of the whole nonnegative numbers. A set I of the transfer airports and a set J of the final destinations are given. For each transfer airport we know a set of the final destinations that can be reached from the given transfer airport. We plan the air routes in a planning time span which is divided into several time periods; a set K of these time periods is given.

For each transfer airport $i \in I$ and each time period $k \in K$ the demand for transport r_{ik} is known. Analogously, for each final destination $j \in J$ and each time period $k \in K$ the demand for transport d_{jk} that will be satisfied if a flight will be planned from the regional airport in the time period $k \in K$ is given. Our task is to decide about how to choose maximally p transfer airports so that the number of the passengers flying from the regional airport is maximized.

It is assumed in the model that each final destination can be reached from at least one transfer airport in at least one time period. Further we presume that at least one final destination is reachable from more than one transfer airport. If we permitted the flights to several transfer airports within single time period, it would probably cause a break-up of single passenger flow into several partial passenger flows. This action results in an increase of operational costs. Therefore we will assume that in each time period only single flight can be planned. Firstly we will present the linear mathematical model that will be consequently fuzzificated. The linear mathematical model for solving the task has the following form:

$$\max\ f(w,y,z) = \sum_{i \in I} \sum_{k \in K} r_{ik}\ z_{ik} + \sum_{j \in J} \sum_{k \in K} d_{jk}\ y_{jk} \tag{1}$$

subject to

$$\sum_{i \in I} w_i \leq p \tag{2}$$

$$\sum_{i \in I} z_{ik} \leq 1 \text{ for } k \in K \tag{3}$$

$$\sum_{k \in K} z_{ik} \leq w_i\ M \text{ for } i \in I \tag{4}$$

$$y_{jk} \leq \sum_{i \in I_{jk}} z_{ik} \text{ for } j \in J \text{ and } k \in K \tag{5}$$

$$w_i \in \{0;1\} \text{ for } i \in I \tag{6}$$

$$z_{ik} \in \{0;1\} \text{ for } i \in I \text{ and } k \in K \tag{7}$$

$$y_{jk} \in \{0;1\} \text{ for } j \in J \text{ and } k \in K \tag{8}$$

Function (1) represents the total satisfied demand of the passengers for transport to the transfer airports and the final destinations. Constraint (2) ensures that we will not operate the flights to more than *p* airports. The group of constraints (3) assures that only single flight can be planned in each time period. The groups of constraints (4) and (5) create logical links among the variables of the model and the groups of constraints (6) – (8) represent domains of definition of the variables used in the model. The symbol *M* corresponds to a so-called prohibitive constant. Its minimal value can be estimated using for example the formula $M = |K|$. The set I_{jk} represents the set of the transfer airports from which the final destination $j \in J$ is reachable in the time period $k \in K$.

It is obvious from the task that uncertainty shows in the demand for transport. The coefficients of the demand are found only in the objective function of the model. From the point of view of modelling uncertainty (according to [4]) we choose the approach that is intended for fuzzification of the objective function's coefficients and to model uncertainty we employ triangle fuzzy numbers. Therefore it is necessary to change the formulation of the problem – the values of the demand for transport represented by the coefficients r_{ik} and d_{jk} must be re-defined. The original coefficients are replaced by the triangle fuzzy numbers $\tilde{r}_{ik} = \langle r_{ik1}; r_{ik2}; r_{ik3} \rangle$ and $\tilde{d}_{jk} = \langle d_{jk1}; d_{jk2}; d_{jk3} \rangle$. Applying the rules discussed in the book [4] we get a new objective function having the following form (9):

$$\max f(w,y,z,h) = \sum_{i \in I} \sum_{k \in K} r_{ik3} z_{ik} + \sum_{j \in J} \sum_{k \in K} d_{jk3} y_{jk} - \left[\sum_{i \in I} \sum_{k \in K} r_{ik3} z_{ik} - \sum_{i \in I} \sum_{k \in K} r_{ik2} z_{ik} + \sum_{j \in J} \sum_{k \in K} d_{jk3} y_{jk} - \sum_{j \in J} \sum_{k \in K} d_{jk2} y_{jk} \right] h . \tag{9}$$

Please note that the symbol *h* denotes the level of satisfaction. Constraints (2) up to (8) are the same in the modified fuzzy model. According to [4] the model (2) – (9) is solved for the pre-defined levels of satisfaction.

4 The computational experiments

The computational experiments were carried out using a model example. The set of the transfer airports has two elements and the set of the final destinations contains 5 elements. The planning time span for which we would like to plan the flights is divided into 2 time periods; that means we plan two flights from the regional airport. The demand for transport to the transfer airports and the final destinations in the individual time periods is defined in Table 1.

The time period	1	2
1st transfer airport	100	130
2nd transfer airport	60	200
1st final destination	10	20
2nd final destination	30	40
3rd final destination	50	30
4th final destination	20	10
5th final destination	5	5

Table 1 The demand for transport (the number of the passengers per a time period).

Table 2 summarizes information about reachability of the final destinations from the transfer airports in the individual time periods.

1st time period			2nd time period		
The transfer airport	1	2	The transfer airport	1	2
1st final destination	1	1	1st final destination	0	1
2nd final destination	0	1	2nd final destination	0	1
3rd final destination	1	0	3rd final destination	1	0
4th final destination	0	1	4th final destination	0	1
5th final destination	0	0	5th final destination	0	1

Table 2 Reachability of the final destinations from the transfer airports.

In our example the regional airport is interested in to operate one flight in each time period and all the flights are to the same transfer airport. The computational experiments were carried out in two phases. In the frame of the first phase the experiments with model (1) – (8) were carried out and the second phase is represented by the experiments with model (2) – (9).

The results of the first phase are depicted in Figure 1. The node v_0 represents the regional airport, the nodes $\overline{v_1}$, $\overline{v_2}$ are the transfer airports and the nodes $\overline{\overline{v_1}},\overline{\overline{v_2}},...,\overline{\overline{v_n}}$ correspond to the final destinations that are reachable from the transfer airports without any transfer. The dashed undirected edges represent reachability the individual transfer airports and the final destinations. The dashed undirected edge is replaced by the bold directed edge if it is made possible to fly from the regional airport to the transfer airport or to the final destination with only one transfer; the bold directed edges represent the demand for transport that is satisfied. The value of the objective function (the demand for transport that is satisfied) is equal to 395 passengers after the first phase. The total number of the passengers in both time periods who leave the system at 2nd transfer airport (they do not fly to any final destination because the transfer airport is their final destination) is equal to 260 and 135 passengers fly to the requested final destinations. The weights of the directed edges in Figure 1 correspond to the number of the passengers who fly in the corresponding relation and in the given time period.

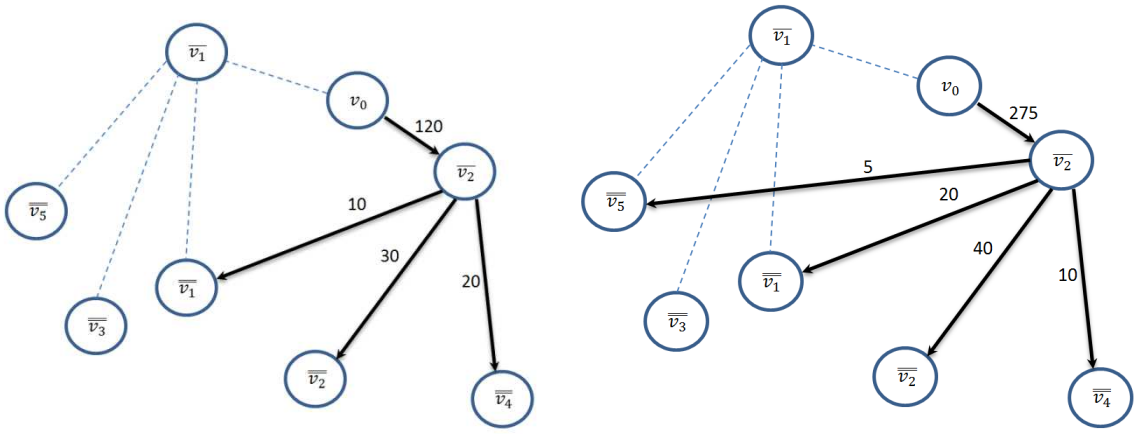


Figure 1 The solution for the first time period (on the left) and for the second time period (on the right).

Let us describe the second phase of the experiment and its solution. Let us assume that the demand for transport per a time period is defined by the triangle fuzzy numbers $\tilde{r}_{ik}$ and $\tilde{d}_{jk}$; it holds for the fuzzy numbers that $\tilde{r}_{ik} = \langle r_{ik2} - 0.2\,r_{ik2}; r_{ik2}; r_{ik2} + 0.2\,r_{ik2} \rangle$ and $\tilde{d}_{jk} = \langle d_{jk2} - 0.2\,d_{jk2}; d_{jk2}; d_{jk2} + 0.2 d_{jk2} \rangle$. The values of the objective function, that we got using the model (2) – (9), are summarized in Table 3 for the pre-defined levels of satisfaction (the values h are graded by $\Delta h = 0.1$). Using the right values of the fuzzy numbers we computed an optimistic estimation of the objective function value – 474 passengers (the optimistic estimation corresponds to the level of satisfaction equal to $h = 0$).

The level of satisfaction	The satisfied demand	The level of satisfaction	The satisfied demand
0.1	466.1	0.6	426.6
0.2	458.2	0.7	418.7
0.3	450.3	0.8	410.8
0.4	442.4	0.9	402.9
0.5	434.5	1.0	395.0

Table 3 The values of the objective function on the pre-defined levels of satisfaction.

The experiments confirmed an expected trend – the number of the checked-in passengers decreases with the increasing level of satisfaction. We can state on the basis of model (2) – (9) that the value of the objective function increases from 395 passengers (the maximal value of the level of satisfaction) up to 475 passengers (the minimal value of the level of satisfaction). The results including the numbers of the checked-in passengers on the pre-defined levels of satisfaction are listed in Table 4.

h	0.0		0.1		0.2		0.3		0.4		0.5	
The time period	1	2	1	2	1	2	1	2	1	2	1	2
1st transfer airport	0	0	0	0	0	0	0	0	0	0	0	0
2nd transfer airport	72	240	70.8	236	69.6	232	68.4	228	67.2	224	66	220
1st final destination	12	24	11.8	23.6	11.6	23.2	11.4	22.8	11.2	22.4	11	22
2nd final destination	36	48	35.4	47.2	34.8	46.4	34.2	45.6	33.6	44.8	33	44
3rd final destination	0	0	0	0	0	0	0	0	0	0	0	0
4th final destination	24	12	23.6	11.8	23.2	11.6	22.8	11.4	22.4	11.2	22	11
5th final destination	0	6	0	5.9		5.8	0	5.7	0	5.6	0	5.5
Σ	474		466.1		458.2		450.3		442.4		434.5	
h	0.6		0.7		0.8		0.9		1.0			
The time period	1	2	1	2	1	2	1	2	1	2		
1st transfer airport	0	0	0	0	0	0	0	0	0	0		
2nd transfer airport	64.8	216	63.6	212	62.4	208	61.2	204	60	200		
1st final destination	10.8	21.6	10.6	21.2	10.4	20.8	10.2	20.4	10	20		
2nd final destination	32.4	43.2	31.8	42.4	31.2	41.6	30.6	40.8	30	40		
3rd final destination	0	0	0	0	0	0	0	0	0	0		
4th final destination	21.6	10.8	21.2	10.6	20.8	10.4	20.4	10.2	20	10		
5th final destination	0	5.4	0	5.3	0	5.2	0	5.1	0	5		
Σ	426.6		418.7		410.8		402.9		395			

Table 4 The satisfied demand for transport (on the corresponding level of satisfaction).

Please note that the experiments were carried out on a personal computer with the processor INTEL® Core™2 Duo CPU E8400 and 3.25 GB of RAM. Calculation times necessary for solving the model were negligible.

5 Conclusions

The paper is focused on the problem how to effectively choose the air routes that should be operated by the regional airport. It is quite common for the regional airport that several passenger flows with lower intensities enter or exit the regional airport. It is not profitable to operate the air routes to all of the final airports; the regional airport operates only selected air routes to some important airports. These airports are transfer airports; the passengers can fly from the transfer airports to their final destinations. For the transfer airports the demand for transport can be also given.

The paper demonstrates the approach which can be used for solving the problem. The paper contains two original mathematical models that are tested on the model example. The first model is suitable for the situations if the demand for transport is exactly given. The second one works with uncertainty in the demand for transport which is defined using the triangle fuzzy numbers. For the model with uncertainty in the demand for transport some optimization computations were carried out on the pre-defined levels of satisfaction.

Acknowledgements

The publication of this contribution has been supported by the project No: CZ.1.07/2.4.00/31.0162 – "Practical competence and professional qualifications improvement in the area of technical education" that is co-financed by the European Social Fund and the state budget of the Czech Republic and by the project SP2013/55 "Research into classical and unconventional models and optimization approaches for increasing the efficiency of transportation and logistics service systems"

References

[1] GOPALAN, R., TALLURI, K. T.: *Mathematical models in airline schedule planning: A survey*. Annals of Operation Research. Baltzer Science Publishers, Amsterdam, 1998, No. 76, pp. 155 – 185.

[2] JANÁČEK, J., JANÁČKOVÁ, M., SZENDREYOVÁ, A., GÁBRIŠOVÁ, L., KOHÁNI, M., JÁNOŠÍKOVÁ, L.: *Navrhovanie územne rozľahlých obslužných systémov*. University of Žilina, Žilina, 2010.

[3] RAMÍK, J., VLACH, M.: *Generalized concavity in optimization and decision making*. Kluwer Publ. Comp., Boston, 2001.

[4] TEODOROVIĆ, D., VUKADINOVIĆ, K.: Traffic Control and transport Planning: A Fuzzy Sets and Neural Networks Approach. Kluwer Publ. Comp., Boston, 1998.

[5] TEICHMANN, D.: *Optimalizace nabídky destinací v podmínkách regionálního mezinárodního letiště*. Perner´s Contacts. Pardubice: University of Pardubice, 2012, No. 27, pp. 211 – 221.

[6] TRANI, A., A., BAIK, H.: *Aplications of linear programming*. In Proceedings of the seminar "Analysis of Air Transportation Systems", Virginia, 2003, June 9 – 12.

[7] ŽIHLA, Z.: *Provozování podniků letecké dopravy a letišť*. CERM, Brno, 2010.

Comparison of market risk models with respect to suggested changes of Basel Accord

Tomáš Tichý[1]

Abstract. The market risk capital charge of financial institutions is mostly calculated by internal models based on integrated Value at Risk (VaR) approach since introduction of the Amendment to Basel Accord in 1996. The internal models should fulfil several quantitative and qualitative criteria. Besides others, it is a so called backtesting procedure, which was one of the main reasons while alternative approach to market risk estimation – conditional Value at Risk or Expected Shortfall (ES) – were not applicable for the purpose of capital charge calculation. However, it is supposed that this approach will be incorporated into Basel III. In this paper we provide an extensive simulation study using various sets of market data to show potential impact of ES on capital requirements.

Keywords: market risk, capital requirements, Value at Risk, Expected shortfall

JEL classification: C44
AMS classification: 90C15

1 Introduction

Market risk capital requirements have been effective as the amendment to the Basel Accord since 1996. Market risk constitute an important part of capital especially at those financial institutions that are internationally active. Several studies (see eg. Berkowitz and O'Brien, 2002) documented that a majority of banks uses historical simulation to Value at Risk in order to estimate the risk exposure. Simultaneously, however, it was documented that historical simulation in its basic nature does not allow to capture the risk well, since it would need very long time series to work better – very long time series, however, would lead to overlooking of structure changes. Another popular approach, a so called variance-covariance approach, which is based on analytical calculation, might be problematic to apply if several distinct parametric models are considered as risk sources.

Very recently Basel committee on banking supervision (BIS, 2013) started to consider replacing Value at Risk by Expected shortfall (ES) with intention to capture a *tail risk*. The idea was to calculate ES for 97.5%, while for backtesting purposes and zone determination a VaR measure was kept (assuming the confidence of 99%). However, it is simultaneously assumed that various holding periods will be assumed for particular broad risk classes, ranging from 10 days (some equity prices) to one year (CDS).

In this paper we focus on FX rate position (ie. new Basel proposal assumes holding period of 20 days) and calculate both VaR and ES for several probabilities in both tails (for example, a long position in a particular FX rate in one bank might imply a short position in another bank). Our aim is to compare the impact of various (overlapping) intervals used for parameter estimation assuming normal distribution and two subordinated Lévy processes (VG, NIG). We proceed as follows – in section 2 and 3 the basic foundations of risk management and models we are going to apply are provided (in line with Tichý, 2011). Next, selected data set are described and risk measures are calculated and compared.

[1]VSB-TU Ostrava, Faculty of Economics, Department of Finance, Sokolska 33, 701 21 Ostrava, Czech Republic, tomas.tichy@vsb.cz

2 Risk management in financial institutions

The two most important motives for managing the risk of financial institutions are (i) *the supervisor's requirements* and (ii) *the equity holders interest*. While supervisors formulate the rules, which must be followed by any entity desiring to do a business in a given industry, the equity holders have an inherent right to determine the policy as the owners. Notwithstanding, due to the specificity of financial institutions and strongly leveraged equity, both, supervisors and equity holders, must based their policy on the interest of potential clients – debtholders.

Obviously, the supervisor approach might be regarded as macroeconomical, since the primary interest is a health of the financial sector and potentially the whole economy, whereas the equity holders may target to a particular mix of stakeholders. In both cases, the key rule is that the capital should be adequate to the riskiness of the portfolio (business activities). However, while supervisors rather state minimal capital requirements, the equity holders opt for its optimal level.

Supervisors approach. The core motive of supervisors' activities in financial industry is to keep it as healthy as possible. Clearly, the organism can be regarded as healthy only if all its components are healthy too. The basic interest of supervisors is therefore to set criteria for healthy financial institutions.

The supervisors policy affects risk management activities of financial institutions in two important ways: (i) first, they specify eligible approaches to risk measuring; (ii) second, they set risk limits, which should not be broken. In banking industry, starting with The Amendment 1992, the financial institutions are eligible to use VaR-based approaches to quantify market risk they are exposed to. The horizon, for which the risk should be monitored is set to ten days[1] and pre-set confidence is 99%.

Assuming a random variable X following a Gaussian distribution, VaR over a time of length Δt at confidence level α (ie. on a probability level $p = 1 - \alpha$) can be obtained as follows:

$$VaR_X(\Delta t, \alpha) = -F_X^{-1}(1-\alpha) = -\mu_X(\Delta t) - \sigma_X(\Delta t) F_{\mathcal{N}}^{-1}(1-\alpha). \tag{1}$$

Here, with some simplification, $-F_X^{-1}(p)$ denotes the inverse function to the distribution function of random variable X for p, which is further decomposed into the mean (the expected value) of random variable X over Δt, $\mu_X(\Delta t)$, and the product of its standard deviation, $\sigma_X(\Delta t)$, and $F_{\mathcal{N}}(p)$ – p-th percentile of standard normal distribution (Gaussian distribution with zero mean and unit variance). Obviously, the formula should be adjusted adequately if different distribution function is considered.

Since the purpose of VaR is to provide the boundary, which is not to be exceeded with some confidence α, there is no statement what happens with probability $p = 1 - \alpha$, ie. when the boundary is broken. For that reason, the conditional value at risk (cVaR) or expected shortfall (ES) was introduced (see eg. Rockafeller and Uraysev, 2002) – the expected result if VaR is exceeded:

$$cVaR_X(\Delta t, \alpha) = -\mathbb{E}[x | x < -VaR(\Delta t, \alpha)]. \tag{2}$$

3 Selected models for returns modelling

We decided to apply three kinds of models, which we denote as follows: BM (Brownian Motion), VG (Variance gamma), NIG (Normal Inverse Gaussian); each of them is evaluated by applying plain Monte Carlo simulation.

BM. The most simple (or standard) model is a (geometric) Brownian motion, since it is based on normal (Gaussian) distribution. That is, as the process driving the innovations a Wiener process with mean value and variance based on standard normal distribution $\mathcal{N}(0,1)$ with probability density function $f_{\mathcal{N}}(x)$ defined as follows:

$$f_{\mathcal{N}}(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \tag{3}$$

is assumed.

[1] As a principle, the market risk is quantified for trading book – the portfolio of liquid and tradable instruments, for which quick rebalancing of positions is feasible.

The main drawback of BM is probably the fact that it does not allow to fit excess kurtosis and skewness. A popular way how to overcome this issue is to adopt a (geometric) Brownian motion driven by a non-decreasing process – such group of models is called subordinated Lévy models. The two models defined below are quite popular in finance and in both cases, the internal process (the subordinator) can be interpreted as the measure of market activity.

VG. The VG model (*variance gamma model*) is one of the most frequently applied multiparametric models of the Lévy family of processes. There exists two ways how to define it – the first one follows the definition of the Brownian motion driven by gamma process.

The probability density function of the gamma process from gamma distribution $\mathcal{G}(a,b)$ with $a = 1/\nu$ and $b = \nu$ is given by:

$$f_{\mathcal{G}}(g,t;\nu) = \frac{g^{\frac{t}{\nu}-1}\exp(-\frac{g}{\nu})}{\nu^{\frac{t}{\nu}}\Gamma(\frac{t}{\nu})}. \tag{4}$$

Since zero mean VG process $\mathcal{VG}(g(t;\nu);\theta,\vartheta)$ can be defined as

$$\mathcal{VG}_t = \theta(g_t - t) + \vartheta\mathcal{Z}(g_t) = \theta(g_t - t) + \vartheta\sqrt{g_t}\varepsilon, \tag{5}$$

where $\varepsilon \in N(0,1)$ and $g \in \mathcal{G}(a,b)$, the VG density function is given as follows:

$$f_{\mathcal{VG}}(x,g(t;\nu);\theta,\vartheta) = \int_0^\infty \frac{1}{\sigma\sqrt{2\pi g}}\exp\left(-\frac{(x-\theta g)^2}{2\vartheta^2 g}\right)\frac{g^{\frac{t}{\nu}-1}\exp(-\frac{g}{\nu})}{\nu^{\frac{t}{\nu}}\Gamma(\frac{t}{\nu})}\,\mathrm{d}g. \tag{6}$$

Thus, we can also get the distribution function $F_{\mathcal{VG}}(x,g(t;\nu);\theta,\vartheta)$:

$$\int_{-\infty}^{x}\int_0^\infty \frac{1}{\sigma\sqrt{2\pi g}}\exp\left(-\frac{(z-\theta g)^2}{2\vartheta^2 g}\right)\frac{g^{\frac{t}{\nu}-1}\exp(-\frac{g}{\nu})}{\nu^{\frac{t}{\nu}}\Gamma(\frac{t}{\nu})}\,dgdz. \tag{7}$$

Alternatively, also the characteristic function of the process can be utilized:

$$\phi_{\mathcal{VG}}(x,g(t;\nu);\theta,\vartheta) = \left(1 - \imath x\theta\nu + \frac{1}{2}\sigma^2\nu x^2\right)^{-t/\nu}. \tag{8}$$

NIG. Another quite popular model, which allows to fit skewness and kurtosis as well, is NIG model (*normal inverse Gaussian model*). NIG model can be defined by the following characteristic function ($\alpha > 0$, $-\alpha < \beta < \alpha$, $\delta > 0$):

$$\phi_{\mathcal{NIG}}(x,t;\alpha,\beta,\delta) = \exp\left[-t\delta\left(\sqrt{\alpha^2-(\beta+\imath x)^2} - \sqrt{\alpha^2-\beta^2}\right)\right]. \tag{9}$$

Then, the density function is given as follows:

$$f_{\mathcal{NIG}}(x,t;\alpha,\beta,\delta) = \frac{\alpha\delta}{\pi}\exp\left(\delta\sqrt{\alpha^2-\beta^2} + \beta x\right)\frac{K_1\left(\alpha\sqrt{\delta^2+x^2}\right)}{\sqrt{\delta^2+x^2}}, \tag{10}$$

where $K_\lambda(x)$ is modified Bessel function:

$$K_\lambda(x) = \frac{1}{2}\int_0^{-\infty} y^{\lambda-1}\exp\left(-\frac{1}{2}x(y+y^{-1})\right)dy. \tag{11}$$

Alternatively, following the definition of the Brownian motion driven by inverse Gaussian (IG) process, ie. process $\mathcal{I}(t;\nu)$ with drift ν, which at time $\mathcal{I} \sim \mathcal{IG}[t;\nu]$ reaches level t, we get the zero-mean process as follows:

$$\mathcal{NIG}(\mathcal{I}(t;\nu);\theta,\vartheta) = \theta(\mathcal{I}_t - t) + \vartheta\mathcal{Z}(\mathcal{I}_t) = \theta(\mathcal{I}_t - t) + \vartheta\sqrt{\mathcal{I}_t}\varepsilon. \tag{12}$$

In this case we can formulate the characteristic function as follows:

$$\phi_{\mathcal{NIG}}(x;\nu,\theta,\vartheta) = \exp\left[\frac{1}{\nu} - \frac{1}{\nu}\left(\sqrt{1+x^2\vartheta^2\nu - 2\theta\nu\imath}\right)\right], \tag{13}$$

which results into: $\theta = \delta\beta/\sqrt{\alpha^2-\beta^2}$, $\vartheta = \frac{\sqrt{\delta\sqrt{\alpha^2-\beta^2}}}{\sqrt{\alpha-\beta}\sqrt{\alpha+\beta}}$ and $\nu = \left(\delta\sqrt{\alpha^2-\beta^2}\right)^{-1}$. Density function $f_{\mathcal{IG}}(x;\delta,\alpha,\beta)$ can be rewritten as:

$$\frac{\delta}{\sqrt{2\pi}}\exp(\delta\sqrt{\alpha^2-\beta^2})x^{-3/2}\exp\left(-\frac{1}{2}\left(\delta^2x^{-1} + (\alpha^2-\beta^2)x\right)\right). \tag{14}$$

4 Data

We consider evolution of 8 FX rates (CZK, GBP, HUF, IDR, JPY, KRW, PLN, USD) with respect to EUR over preceding 10 years, which gives us 2520 observations of log-returns. Before estimating the parameters of expected return, standard deviation, skewness and kurtosis we normalize the data so that they have zero mean and unite variance for holding periods of 10 days, ie. related 10-day VaR/ES can be easily interpreted (recalculated) on a percentage basis.[2]

Basic descriptive statistics of particular FX rates after normalization and recalculation on two-weeks horizon (ie. times 10 days) are depicted in Table 1. It is apparent that the tails are not symmetric for most of the currencies and that for more developed economy and more liquid pairs the kurtosis is rather lower than for less liquid and efficient (IDR and KRW, but also HUF and PLN).

Next, in Table 2, empirical quantiles (ie. $-VaR$) and ES are provided for both tails. Since all the data were normalized for the mean and standard deviation, all quantiles and ES should be the same if the assumption of normality holds. Obviously, it is not the case – the results are different for each FX rate and the ranking is generally related to the observed kurtosis. Obviously, the differences are higher for ES. Note also, that the new Basel Accord suggests to replace $VaR_{0.01}$ by $ES_{0.025}$ – in the case of FX rates considered here, it would lead to even 10% increase in the capital requirement. Such number would further increase if we admit twice as long holding period. Longer horizon of data collecting often leads to higher skewness, but lower kurtosis and thus, ES shortfall will be lower.

Daily data	*mean*	*median*	*st.deviation*	*skewness*	*kurtosis*
CZK	0	0.018	1	0.0314	10.2736
GBP	0	0.0678	1	−0.4706	8.0269
HUF	0	0.0829	1	−0.6868	14.0017
IDR	0	0.0121	1	0.3571	16.0597
JPY	0	−0.1513	1	0.1799	8.9864
KRW	0	0.0295	1	−0.0726	13.8782
PLN	0	0.1259	1	−0.3377	14.5372
USD	0	−0.0403	1	0.0331	5.8854

Table 1 Descriptive statistics of daily FX log-returns

5 Results

In this section we calculate VaR and ES for four probability levels in both tails $(0.001, 0.01, 0.025, 0.05)$ assuming three kinds of models. In Table 3, results obtained on a daily basis dataset via Monte Carlo simulation with 100000 scenarios are provided.

Normal distribution

Normal distribution considers only first two moments and thus, after normalization, we should obtain the same results for each currency (apparently we see that plain Monte Carlo with 100 000 scenarios is not sufficient to get exactly the same results). We can also see that $ES_{0.025}$ significantly approaches to $VaR_{0.01}$ in both tails as was assumed by Basel III proposal.

VG and NIG distribution

As concerns VG model, we take into account also the higher moments and thus the results for particular FX rates clearly differs. It also seems that $ES_{0.025}$ is not so far to $VaR_{0.01}$ as according to empirical observations – it might be given by different tail shape.

[2]The reason why all the results are recalculated for assumed 10-day holding period is to keep them consistent with results obtained in previous research (elaborated in line with Basel I/II). Due to lack of space only results for daily data are provided. Additional tables with results on 5/10/20-day data can be available upon request.

Quantile	0.001	0.01	0.025	0.05	0.5	0.95	0.975	0.99	0.999
CZK	−5.6143	−2.8755	−2.045	−1.509	0.0057	1.5622	2.0327	2.779	5.8417
GBP	−5.6102	−2.5931	−1.9527	−1.5315	0.0212	1.4891	2.0049	2.5766	4.4636
HUF	−7.0892	−2.6879	−2.0836	−1.5511	0.0259	1.418	1.8174	2.5358	5.3382
IDR	−5.4767	−2.8043	−1.9926	−1.4404	0.0028	1.4524	1.8793	2.6917	6.1415
JPY	−4.6802	−2.4561	−1.8496	−1.4414	−0.0478	1.6121	2.1472	2.9555	4.5777
KRW	−6.724	−2.8242	−1.8781	−1.4828	0.0092	1.4717	1.8367	2.6911	5.6915
PLN	−5.8227	−3.0876	−2.0669	−1.5422	0.0392	1.356	1.7234	2.447	4.2789
USD	−4.0446	−2.5554	−2.0005	−1.5587	−0.013	1.6204	1.9189	2.6639	3.9437
ES	0.001	0.01	0.025	0.05	0.5	0.95	0.975	0.99	0.999
CZK	6.7655	3.8741	3.0132	2.3899	− − −	2.3483	2.9151	3.8531	7.0296
GBP	6.4705	4.0192	2.9927	2.3608	− − −	2.1925	2.6736	3.3101	4.9229
HUF	7.7198	4.6079	3.248	2.4923	− − −	2.1547	2.6939	3.6543	7.2432
IDR	7.6944	3.8119	2.948	2.3226	− − −	2.3071	2.9424	4.1201	7.9819
JPY	6.8955	3.4926	2.6516	2.1399	− − −	2.4012	2.9664	3.7341	6.7558
KRW	7.5872	4.1775	3.053	2.3536	− − −	2.2821	2.8881	3.9706	7.2744
PLN	7.3579	4.2961	3.1941	2.4902	− − −	2.1019	2.6599	3.7311	8.907
USD	6.0425	3.2694	2.6594	2.1933	− − −	2.199	2.6376	3.2837	5.658

Table 2 Quantiles (ie. $-VaR$) and Expected shortfalls for both tails

NIG model is by definition very similar to VG model and although behaviour of their probability functions slightly differ (especially with the sharpness of the peak), the differences in the results might be given by insufficient preciseness of the simulation.

If we compare ES and VaR for probability levels suggested by Basel III, we can see important differences with estimation based on daily dataset. Notwithstanding, in case that we assume the holding period of 20 days (ie. we use such horizon for the estimation), $ES_{0.025}$ can be even lower than $VaR_{0.01}$.

6 Conclusion

Currently it is supposed that ES will replace VaR in measuring risk of financial institutions. In this paper we have empirically shown some inconsistency in the assumed calibration of probability levels for VaR and ES and subsequent recalculation for different horizons, though there is serious impact of the dataset basis we consider. The impact might be even more serious if portfolios will be considered.

Acknowledgements

The research was supported through the Czech Science Foundation (GAR) under project 13-13142S and furthermore by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and the European Social Fund (CZ.1.07/2.3.00/20.0296) and SP2014/16, an SGS research project of VSB-TU Ostrava.

References

[1] Berkowitz, J., O'Brien, J.: How accurate are value-at-risk models at commercial banks? *Journal of Finance* 57(3): 1093–1111, 2002.

[2] Kresta, A., Tichý, T.: International Equity Portfolio Risk Modeling: The case of NIG model and ordinary copula functions. *CJEF* 61 (2): 141–151, 2012.

[3] Rockafeller, R.T., Uraysev, S.: Conditional Value at Risk for general loss distributions. *Journal of Banking and Finance 26*, 1443–1471, 2002.

[4] Tichý, T. *Lévy Processes in Finance: Selected applications with theoretical background.* SAEI, vol. 9. Ostrava: VSB-TU Ostrava, 2011.

Normal distribution									
Quantile	0.001	0.01	0.025	0.05	0.5	0.95	0.975	0.99	0.999
CZK	−3.0696	−2.3296	−1.9584	−1.6425	0.0022	1.6456	1.9571	2.3252	3.0876
GBP	−3.0823	−2.324	−1.9581	−1.6444	−0.0009	1.6441	1.9603	2.3315	3.0927
HUF	−3.0829	−2.3309	−1.9654	−1.6482	−0.0028	1.6447	1.9628	2.3265	3.0699
IDR	−3.0791	−2.3265	−1.9603	−1.6425	0.0013	1.645	1.9612	2.3271	3.0946
JPY	−3.0971	−2.3249	−1.9578	−1.6444	−0.0006	1.645	1.96	2.3246	3.106
KRW	−3.0908	−2.3252	−1.9584	−1.6441	0.0006	1.6415	1.9587	2.3315	3.1107
PLN	−3.0854	−2.3205	−1.9565	−1.6434	0.0003	1.6428	1.9562	2.3246	3.1073
USD	−3.0924	−2.3208	−1.9562	−1.6453	−0.0019	1.6453	1.96	2.3265	3.0892
ES	0.001	0.01	0.025	0.05	0.5	0.95	0.975	0.99	0.999
CZK	3.3379	2.6594	2.3364	2.0608	− − −	2.0619	2.3367	2.6658	3.3709
GBP	3.3661	2.6615	2.3352	2.061	− − −	2.0632	2.339	2.6661	3.3571
HUF	3.3546	2.6618	2.3396	2.0656	− − −	2.0635	2.337	2.6585	3.3615
IDR	3.3572	2.6647	2.338	2.0621	− − −	2.0642	2.3398	2.6688	3.379
JPY	3.3666	2.6661	2.3378	2.0618	− − −	2.0631	2.3391	2.6691	3.3814
KRW	3.3816	2.6651	2.3371	2.0619	− − −	2.0643	2.3433	2.6759	3.3892
PLN	3.3568	2.6579	2.3325	2.059	− − −	2.0607	2.3369	2.6668	3.3845
USD	3.3752	2.6641	2.3349	2.0607	− − −	2.0619	2.3362	2.6626	3.3703
Variance gamma distribution									
Quantile	0.001	0.01	0.025	0.05	0.5	0.95	0.975	0.99	0.999
CZK	−5.3044	−3.038	−2.1896	−1.5742	−0.0041	1.5789	2.1965	3.0538	5.3537
GBP	−5.2497	−3.112	−2.3062	−1.6928	0.0689	1.5147	2.0204	2.7164	4.4952
HUF	−6.1946	−3.4029	−2.3518	−1.6083	0.0632	1.4262	2.0498	2.9248	5.299
IDR	−5.7933	−3.131	−2.1368	−1.4404	−0.0275	1.5226	2.274	3.3685	6.3091
JPY	−4.9325	−2.8979	−2.1238	−1.5634	−0.0256	1.6292	2.231	3.0519	5.2244
KRW	−5.8584	−3.2151	−2.2218	−1.5299	0.0066	1.5065	2.1861	3.1477	5.7601
PLN	−6.1541	−3.3112	−2.2759	−1.5479	0.0294	1.4638	2.1374	3.0987	5.6912
USD	−4.3168	−2.73	−2.0988	−1.6204	−0.006	1.6384	2.1269	2.7724	4.4117
ES	0.001	0.01	0.025	0.05	0.5	0.95	0.975	0.99	0.999
CZK	6.2941	4.002	3.131	2.4879	− − −	2.4992	3.1503	4.0341	6.3101
GBP	6.1762	4.0219	3.1988	2.5802	− − −	2.2605	2.7826	3.4812	5.2837
HUF	7.5414	4.623	3.5237	2.7286	− − −	2.3655	3.0314	3.9515	6.371
IDR	7.0038	4.2745	3.242	2.4935	− − −	2.6705	3.491	4.6322	7.6565
JPY	5.8158	3.7727	2.9786	2.3938	− − −	2.5167	3.1409	3.9867	6.171
KRW	7.0871	4.3519	3.3219	2.5777	− − −	2.5303	3.2597	4.2637	6.8969
PLN	7.4731	4.5359	3.439	2.6526	− − −	2.4826	3.2096	4.2167	6.929
USD	4.997	3.4149	2.7858	2.3084	− − −	2.3433	2.8311	3.4775	5.076
Normal inverse Gaussian distribution									
Quantile	0.001	0.01	0.025	0.05	0.5	0.95	0.975	0.99	0.999
CZK	−5.3224	−2.8906	−2.0662	−1.5128	−0.0016	1.5192	2.0798	2.909	5.3591
GBP	−5.2993	−3.0007	−2.1905	−1.6241	0.0386	1.4885	1.9439	2.5823	4.3832
HUF	−6.3078	−3.1983	−2.1902	−1.5328	0.0367	1.4031	1.9404	2.7449	5.2095
IDR	−5.7908	−2.9033	−1.9916	−1.4025	−0.0168	1.454	2.1016	3.1193	6.3726
JPY	−4.9259	−2.7774	−2.0324	−1.5163	−0.0117	1.5644	2.1263	2.9416	5.3221
KRW	−5.8398	−3.0203	−2.0786	−1.4641	0.0032	1.4531	2.0488	2.9476	5.7164
PLN	−6.2076	−3.1038	−2.1257	−1.4834	0.0155	1.4233	1.9922	2.8799	5.5912
USD	−4.3529	−2.6683	−2.0375	−1.5837	−0.0041	1.6014	2.065	2.7085	4.4376
ES	0.001	0.01	0.025	0.05	0.5	0.95	0.975	0.99	0.999
CZK	6.4945	3.922	3.0123	2.3828	− − −	2.399	3.0349	3.9505	6.5665
GBP	6.3536	3.9875	3.1136	2.4916	− − −	2.1793	2.6713	3.3588	5.2293
HUF	7.9194	4.5257	3.3737	2.5946	− − −	2.2588	2.8806	3.7886	6.3533
IDR	7.1826	4.1152	3.0665	2.361	− − −	2.5191	3.3072	4.4883	8.0432
JPY	5.9286	3.689	2.8822	2.3108	− − −	2.4373	3.065	3.9611	6.5227
KRW	7.2812	4.2276	3.1702	2.4482	− − −	2.4073	3.1034	4.13	7.1103
PLN	7.7355	4.4062	3.2822	2.5203	− − −	2.3462	3.026	4.0317	6.9537
USD	5.1441	3.3953	2.7393	2.2612	− − −	2.2933	2.7819	3.4524	5.2273

Table 3 Estimated Quantiles (ie. $-VaR$) and ES for both tails (ie. short and long position)

Max-prod eigenvectors and their applications in decision making

Hana Tomášková[1], Martin Gavalec[2]

Abstract. The max-prod eigenvectors play very important role in the AHP approach. The number of independent eigenvectors can be considered as the degree of the preference matrix consistency. If the number is greater than one, then there exist more than one consistent approximation of the preference matrix. For the consistent approximation determined by the found independent eigenvectors, their distance from the original preference matrix can be specified. The distance function then can be used to find an optimal consistent approximation. The method is illustrated by numerical examples.

Keywords: preference matrix, tropical algebra, AHP, eigenvectors

JEL classification: C44
AMS classification: 90C15

1 Introduction

The standard Analytic Hierarchy Process (AHP) method formulated by T. Saaty [5] uses an eigenvector of the preference matrix computed by the methods of linear algebra. In the AHP method a structured model of the problem is created, which helps to find from the set of possible alternatives the most suitable one.

The preferences given by human experts are often inconsistent and do not reflect the deep relations between the processed notions, see [4]. The standard approach to the problem of finding the relative importance vector out of an inconsistent preference matrix uses an eigenvector of the preference matrix computed by the methods of linear algebra, [6].

A different approach to inconsistency is described in this paper. The max-prod algebra combining the maximum operation with the product operation of positive real numbers (so-called: tropical algebra) is used, and the relative importance vector is computed in tropical max-prod algebra as an eigenvector of the given preference matrix. A polynomial algorithm for efficient computation of the optimal consistent approximation is described in [2]. The work of the algorithm is shown on examples.

2 Max-prod algebra

Max-prod algebra is one of most frequently used tropical algebras. The choice or this algebra is motivated by the fact that the preference matrices are written in the multiplicative form.

By max-prod algebra $(\mathcal{R}^+, \oplus, \otimes)$ we understand a linear structure on the linearly ordered set $\mathcal{R}^+$ of positive real numbers together with the binary operations $\oplus$ = maximum and $\otimes$ = product, analogously as the ordinary addition and multiplication operations are used in the classical linear algebra. The operations $\oplus$ and $\otimes$ are extended to matrices and vectors over $\mathcal{R}^+$ in the standard way. Notation $\mathcal{R}^+(n,n)$ $(\mathcal{R}^+(n))$ is used below for the set of all square matrices (vectors) of dimension n.

The eigenproblem in max-prod algebra can be formulated as follows. Given matrix $A \in \mathcal{R}^+(n,n)$, find $\lambda \in \mathcal{R}^+$ and $X \in \mathcal{R}^+(n)$ such that

$$A \otimes X = \lambda \otimes X \tag{1}$$

[1]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: hana.tomaskova@uhk.cz
[2]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: martin.gavalec@uhk.cz

The eigenproblem in tropical algebra, as well as other important problems such as linear optimization, have been studied for several decades by many authors. The results from the monograph [1] formulated below will be useful in further sections.

Given matrix $A \in \mathcal{R}^+(n,n)$, notation $G(A)$ is used for the complete directed graph with n nodes, in which the edges are evaluated by the matrix inputs, e.i. the weight of the edge (i,j) is $w(i,j) = a_{ij}$ for every i,j. The set of all cycles in $G(A)$ is denoted by $C(A)$ and $\lambda(A)$ denotes the maximal geometric cycle mean in $G(A)$, i.e.

$$\lambda(A) = \max_{C \in C(A)} \left(\prod_{(i,j) \in C} a_{ij} \right)^{1/|C|} \tag{2}$$

The maximal geometric cycle mean can be computed in $O(n^3)$ time be slightly modified Karp's method, see [3].

Theorem 1. *[1] Let $A \in \mathcal{R}^+(n,n)$, $B = A \otimes \lambda(A)^{-1}$ and $\Gamma(B) = B \oplus B^2 \oplus \cdots \oplus B^{n-1}$. The columns of matrix $\Gamma(B)$ having their diagonal value 1 are eigenvectors of B with eigenvalue value 1 and they are also the eigenvectors of the original matrix A with eigenvalue $\lambda(A)$. Moreover, $\lambda(A)$ is the only eigenvalue of A.*

Matrix $\Gamma(B)$ is called the *weak transitive closure* of B and the eigenvectors described in Theorem 1 are called the *fundamental eigenvectors* of A.

Theorem 2. *[1] Let $A \in \mathcal{R}^+(n,n)$, $x \in \mathcal{R}^+(n)$. Then x is eigenvector of A if and only if x is a max-prod linear combination of the fundamental eigenvectors of A.*

3 Max-prod eigenvectors and consistent preference matrices

Let $A_1, A_2, .., A_n$ be a set of alternatives. We shall use the notation N for the set $\{1, 2, \ldots, n\}$ and $\mathcal{R}^+$ for the set of all positive real numbers. The quantified judgments on pairs A_i, A_j are represented by an $n \times n$ matrix $A = (a_{ij})$, $i,j \in N$. The entries a_{ij} are positive real values and they are interpreted as multiplicative evaluation of the relative preference of the alternative A_i in respect to A_j. That is, the preference of A_i is considered to by the $a_{ij}-$multiple of the preference of A_j. Consequently, the preference of A_j is the $1/a_{ij}-$multiple of the preference of A_j, and $a_{ii} = 1$ for every $i \in N$. In this interpretation, matrix A is called *preference matrix* of the alternatives $A_1, A_2, .., A_n$. The basic properties of preference matrices are defined as follows

A is *unidiagonal* if $a_{ii} = 1$ for every $i \in N$
A is *reciprocal* if $a_{ij} = 1/a_{ji}$ for every $i,j \in N$
A is *consistent* if $a_{ij} \cdot a_{jk} = a_{ik}$ for every $i,j,k \in N$

It is easy to see that if A is reciprocal, then it is unidiagonal, and if A is consistent, then it is reciprocal. Any reciprocal matrix has the form

$$A = \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{pmatrix} \tag{3}$$

In real problems, the prefence matrix is created by an expert.. Standardly, the expert's preference matrix is reciprocal and unidiagonal, because these two properties are easy to verify. On the other hand, the consistency of a matrix cannot be directly seen. Therefore, the preference matrices given by experts are usually inconsistent.

The problem of finding a consistent approximation of a given preference matrix A has been studied in [2], where an algorithm for computing the optimal approximation of the expert's preference matrix has been described.

We say that vectors $x, y \in \mathcal{R}^+(n)$ are *colinear*, if there is $c \in \mathcal{R}^+$ such that $x = c \cdot y$. Consistent matrices were characterized in [2] by the theorems below.

Theorem 3. *[2] Let $A \in \mathcal{R}^+(n,n)$ be a unidiagonal matrix. Then the following assertions are equivalent*

(i) *A is consistent,*
(ii) *every two columns of A are colinear.*

Theorem 4. *[2] Let $A \in \mathcal{R}^+(n,n)$ be a consistent matrix. Then*

(i) *$A \otimes x = x$ for every column x of A.*

4 Optimization algorithm

The optimization algorithm $\mathcal{A}_{CONSIST}$ described below solves this problem: for the input preference matrix A (which may be possibly inconsistent), find a consistent matrix $\tilde{A}$ (called *the optimal consistent approximation* of A), which will be as close to A as possible. See [2] for details .

The algorithm works with two assumptions:
(A1) the input matrix A is unidiagonal and reciprocal,
(A2) the columns of the output matrix $\tilde{A}$ are eigenvectors of A .

Similarly as in Theorem 4, the eigenvectors are considered in max-prod algebra, and the distance $d(A,\tilde{A})$ of matrices A, $\tilde{A}$ will be computed as the norm $\|D\|$

$$d(A,\tilde{A}) = \|D\| = \bigoplus_{i,j\in N} d_{ij} \quad \text{with} \quad d_{ij} = \frac{a_{ij}}{\tilde{a}_{ij}} \tag{4}$$

It is easy to see that $d_{ij} = 1/d_{ji}$ for every $i,j \in N$, by the reciprocity of A and $\tilde{A}$. Therefore, the distance function defined in (4) is symetrical. Moreover, $d(A,\tilde{A}) \geq 1$, and $d(A,\tilde{A}) = 1$ if and only if $A = \tilde{A}$ (we recall that 1 is the neutral element in max-prod algebra).

$\mathcal{A}_{CONSIST}$

1 input – a unidiagonal and reciprocal matrix A ;

2 find the eigenvalue $\lambda = \lambda(A)$ of matrix A ;

3 put $B := A \otimes \lambda^{-1}$;

4 compute the weak transitive closure $\Gamma(B) = B \oplus B^2 \oplus \cdots \oplus B^{n-1}$;

5 denote as $u_1, u_2, \ldots, u_h$ the columns of $\Gamma(B)$ containing value 1 on the diagonal (fundamental eigenvectors) ;

6 find values $\beta_1, \beta_2, \ldots, \beta_h$, vector $w = \bigoplus_{k=1}^{h} \beta_k \otimes u_k$ and the consistent matrix W such that every column of W is a multiple of w and the distance $d(A,W)$ is minimal ;

7 output – the optimal consistent approximation $\tilde{A} = W$.

The optimization problem solved in 6 can be formulated as follows. Let $A \in \mathcal{R}^+(n,n)$ be a unidiagonal and reciprocal matrix, let $u, v \in \mathcal{R}^+(n)$ be eigenvectors of A. For any $\beta, \gamma \in \mathcal{R}^+$ denote

$$w(\beta,\gamma) = \beta \otimes u \oplus \gamma \otimes v \ , \tag{5}$$

and define a unidiagonal matrix $W(\beta,\gamma)$ whose columns are multiples of vector $w(\beta,\gamma)$. The optimization problem consists of finding such values β, γ, for which the distance $d(A, W(\beta,\gamma))$ is minimal. That is,

$$d(A, W(\beta,\gamma)) = \bigoplus_{i,j\in N} \frac{w_{ij}(\beta,\gamma)}{a_{ij}} \longrightarrow \min \tag{6}$$

Two equivalent formulations of the problem are: find minimal value $\varepsilon \geq 1$ such that there are $\beta, \gamma \in \mathcal{R}^+$ such that

$$\bigoplus_{i,j\in N} \frac{\big(\beta \otimes u_i \oplus \gamma \otimes v_i\big)}{a_{ij} \otimes \big(\beta \otimes u_j \oplus \gamma \otimes v_j\big)} \leq \varepsilon \ , \tag{7}$$

or: find the minimal value $\varepsilon \geq 1$ for which the system of inequalities

$$\bigoplus_{i \in N} \frac{\left(\beta \otimes u_i \oplus \gamma \otimes v_i\right)}{a_{ij}} \leq \left(\beta \otimes u_j \oplus \gamma \otimes v_j\right) \otimes \varepsilon \ , \tag{8}$$

with $j \in N$, is solvable in variables β, γ.

5 Numerical examples

The work of algorithm $\mathcal{A}_{CONSIST}$ is illustrated by an examples. First example illustrate the calculation of AHP decision for the purpose of choosing a stock title. Second example illustrate the problem of inconsistent matrix, which consistent ratio in classical linear algebra is unsuitable and we can optimize the solution.

Example 1. Example is illustrate on the problem of choosing a stock title. Matrix A contains preference matrix for 7 criteria (for example: Dividend Yield, EPS, Sale, Payout Ratio, Return on Equity, P/E, Price).

$$A = \begin{pmatrix} 1 & 1/6 & 1/4 & 2 & 1/5 & 4 & 5 \\ 6 & 1 & 3 & 7 & 2 & 9 & 9 \\ 4 & 1/3 & 1 & 5 & 1/2 & 7 & 8 \\ 1/2 & 1/7 & 1/5 & 1 & 1/6 & 3 & 4 \\ 5 & 1/2 & 2 & 6 & 1 & 8 & 9 \\ 1/4 & 1/9 & 1/7 & 1/3 & 1/8 & 1 & 2 \\ 1/5 & 1/9 & 1/8 & 1/4 & 1/9 & 1/2 & 1 \end{pmatrix}$$

The maximal geometric cycle mean $\lambda(A) = 1.607$ found by the Karp method is the only eigenvalue of A. Then $\mathcal{A}_{CONSIST}$ computes matrix $B = A \otimes \lambda(A)^{-1}$ and the weak transitive closure $\Gamma(B) = B \oplus B^2 \oplus \cdots \oplus B^{n-1}$.

$$\Gamma(B) = \begin{pmatrix} 1 & 0.215 & 0.402 & 1.250 & 0.268 & 2.489 & 3.112 \\ 4.647 & 1 & 1.867 & 5.809 & 1.245 & 11.568 & 14.461 \\ 2.489 & 0.536 & 1 & 3.112 & 0.667 & 6.196 & 7.746 \\ 0.8 & 0.172 & 0.321 & 1 & 0.214 & 1.991 & 2.489 \\ 3.112 & 0.669 & 1.25 & 3.889 & 0.833 & 7.746 & 9.682 \\ 0.4 & 0.86 & 0.161 & 0.5 & 0.107 & 0.996 & 1.245 \\ 0.321 & 0.069 & 0.129 & 0.402 & 0.089 & 0.8 & 1 \end{pmatrix}$$

The fundamental eigenvectors of A are the columns of matrix $\Gamma(B)$ with entries 1 on the diagonal, i.e. all columns with the exception of columns 5 and 6. It is easy to verify that columns 1, 2, 3, 4 and 7 are multiples of vector u. Vector v represents a solution of linear calculation of preferences is SW Criterium Decision Plus.

$$u = \begin{pmatrix} 0.078 \\ 0.364 \\ 0.195 \\ 0.063 \\ 0.244 \\ 0.031 \\ 0.025 \end{pmatrix} \text{ confrontation } \quad v = \begin{pmatrix} 0.076 \\ 0.37 \\ 0.187 \\ 0.054 \\ 0.263 \\ 0.028 \\ 0.021 \end{pmatrix} .$$

Example 2. Larger inconsistent preference matrix could have more than one eigenvector. In this case

is important to find an optimal solution with minimal distance between.

$$A = \begin{pmatrix} 1 & 8 & 0.2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 0.125 & 1 & 8 & 2 & 2 & 2 & 2 & 2 & 2 \\ 5 & 0.125 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ 0.5 & 0.5 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \\ 0.5 & 0.5 & 0.5 & 1 & 1 & 1 & 2 & 2 & 2 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 & 1 & 1 & 2 & 2 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 1 & 1 & 8 & 0.2 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.125 & 1 & 8 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 5 & 0.125 & 1 \end{pmatrix}$$

The maximal geometric cycle mean $\lambda(A) = 6.839904$ found by the Karp method is the only eigenvalue of A. Then $\mathcal{A}_{CONSIST}$ computes matrix $B = A \otimes \lambda(A)^{-1}$ and the weak transitive closure $\Gamma(B) = B \oplus B^2 \oplus \cdots \oplus B^{n-1}$.

$$\Gamma(B) = \begin{pmatrix} 1 & 1.170 & 1.368 & 0.342 & 0.4 & 0.4 & 0.4 & 0.468 & 0.547 \\ 0.855 & 1 & 1.170 & 0.292 & 0.342 & 0.342 & 0.342 & 0.4 & 0.468 \\ 0.731 & 0.855 & 1 & 0.250 & 0.292 & 0.292 & 0.292 & 0.342 & 0.4 \\ 0.107 & 0.125 & 0.146 & 0.146 & 0.146 & 0.292 & 0.292 & 0.342 & 0.4 \\ 0.073 & 0.085 & 0.1 & 0.146 & 0.146 & 0.146 & 0.292 & 0.342 & 0.4 \\ 0.073 & 0.085 & 0.1 & 0.073 & 0.146 & 0.146 & 0.250 & 0.292 & 0.342 \\ 0.1 & 0.117 & 0.137 & 0.1 & 0.1 & 0.146 & 1 & 1.170 & 1.368 \\ 0.085 & 0.1 & 0.117 & 0.085 & 0.085 & 0.125 & 0.855 & 1 & 1.170 \\ 0.073 & 0.085 & 0.1 & 0.073 & 0.073 & 0.107 & 0.731 & 0.855 & 1 \end{pmatrix}$$

The fundamental eigenvectors of A are the columns of matrix $\Gamma(B)$ with entries 1 on the diagonal, i.e. all columns with the exception of columns 4,5,6. It is easy to verify that columns 1, 2, 3 are multiples of vector u, while columns 7,8,9 are multiples of v, where

$$u = \begin{pmatrix} 0.323 \\ 0.276 \\ 0.236 \\ 0.035 \\ 0.024 \\ 0.024 \\ 0.032 \\ 0.028 \\ 0.024 \end{pmatrix}, \quad v = \begin{pmatrix} 0.090 \\ 0.077 \\ 0.066 \\ 0.066 \\ 0.066 \\ 0.056 \\ 0.224 \\ 0.192 \\ 0.164 \end{pmatrix}.$$

Therefore, all eigenvectors of A are max-prod linear combinations of vectors u, v. The optimal solution is found by the optimization process described in this paper. In the above notation, the computation gives the the optimal solution for parameters thats are $0 \leq \frac{\beta}{\gamma} \leq 0.069$.

6 Conclusions

New approach to the inconsistency problem of preference matrices has been described. The suggested solution is based on the tropical max-prod algebra. Using the maximum and product operations on the set of positive real numbers, extended to matrices and vectors, the consistency of a preference matrix has been equivalently characterized by the property that all columns of the matrix are its eigenvectors in max-prod algebra, if all diagonal elements are equal to 1.

Coming out of this fact, it is natural to search for the optimal consistent approximation of a given (possibly inconsistent) preference matrix A, in the set of all consistent matrices the columns of which are

max-prod eigenvectors of A. All these eigenvectors are max-prod linear combinations of finitely many fundamental eigenvectors, and the question arises how to find the optimal combination which corresponds to the consistent approximation with the minimal distance from A.

Polynomial algorithm $\mathcal{A}_{CONSIST}$ for the above optimization problem has been presented and its correctness has been demonstrated. The work of the algorithm is shown on numerical examples.

Acknowledgements

The support of the Czech Science Foundation project # 14-02424S is gratefully acknowledged.

References

[1] Butkovič, P. (2010). *Max-linear Systems: Theory and Algorithms*, Springer Monographs in Mathematics, Springer-Verlag, 2010. .

[2] M. Gavalec, Tomášková, *Consistent approximation of preference matrices and max-prod eigenvectors*, Central European Journal of Operations Research (submitted) .

[3] Karp, R.M. (1978). *A characterization of the minimum cycle mean in a digraph*, Discrete Math. **23** (1978), 309–311.

[4] Ramík, J., Perzina, R. (2010). *A method for solving fuzzy multicriteria decision problems with dependent criteria*, Fuzzy Optimization and Decision Making, Springer Verlag,Vol.9, No.2, 123-141.

[5] Saaty, T.L. (1980). *The analytic hierarchy process*. McGraw-Hill, New York.

[6] Saaty, T.L. (2003). *Decision-making with the AHP: Why is the principal eigenvector necessary*, European Journal of Operational Research, Volume 145, Issue 1, 16 February 2003, Pages 85-91.

Analysis of the Interconnection of the Czech and Global Stock Markets

Filip Tošenovský[1], Elena Mielcová[2]

Abstract. The aim of the paper is to analyze relations among major stock indices. The indices include the Dow Jones and S&P 500 indices from the U.S., the Japan's Nikkei 225 index, the Germany's DAX index and the Czech index PX-50. The authors of the paper are interested, among other things, in mutual dependence of the indices and causal directions of their effects. To analyze these problems, historical time series of the indices are used and scrutinized for presence of nonstationarities potentially present in the series. After analyzing the nonstationarities, a multivariate time series model is built, and its parameters are estimated. The model is used for general conclusions regarding mutual interconnections among the major stock markets. One of the objectives of the paper is to confirm or reject the general expectation related to what extent and with what delay the Prague Stock Exchange responds to developments in other financial markets.

Keywords: stock, index, dependence, time series.

JEL Classification: C22
AMS Classification: 62P20

1 Introduction

Stock markets are commonly believed to react sensitively to economic news. Experience seems to support the view that stock prices are influenced by a wide variety of unanticipated events [3]. One such event can be the effect of dynamics of another stock market - a stock market from another country, for example. This influence can be even stronger when comparing a minor capital market with major markets. To reveal such an influence, this paper is devoted to the idea of confirming or rejecting the general expectation related to what extent and with what delay the Prague Stock Exchange, as a minor stock market, responds to developments in other financial markets. In order to reveal the mutual relationships between the Prague Stock Exchange and major stock exchanges, daily closing levels of the Dow Jones and S&P 500 indices from the U.S., the Japan's Nikkei 225 index, the Germany's DAX index and the Czech index PX-50 were selected.

As a tool to compare the development trends, we considered multivariate time-series models applied to time series of the selected stock exchange indices – namely, a vector autoregression model (VAR) and a vector error-correction model (VECM). However, comparisons of the main stock market indices revealed problems with multicollinearity, and this led us to using a bivariate analysis of the PX-50 and all other indices. The existence of long-term dynamics was tested using Engle-Granger test; a short-term dynamics was studied by error-correction model [16]. Similar approach is described in Alexander (1999) [1], who explained the role of cointegration analysis in financial markets as an advanced tool for dealing with financial data. He reviewed Engle-Granger methodology, construction of the error-correction model and explained how the model captures dynamic correlations and causalities between returns of asset prices. Long-run and short run dynamics were studied on different kinds of data - in economics, for example, Batini and Greenslade (2006) [2] studied short-run dynamics of non-accelerating inflation rate of unemployment (NAIRU), Kim (1990) [10] estimated purchasing power parity and Miller (1991) [13] studied dynamics of long-run and short-run money demands; Goyenko and Ukhov (2009) [7] revealed liquidity linkage between stock and treasury bond markets using Granger causalities. Studies related to stock market cover, for example, stock price-volume comparison [9], stock prices and dividends relation [14], dynamics of stock index and its future returns [15], comparison of linked security markets [8], volatilities [11] or stock index future trading on returns [6].

[1] Silesian University in Opava/School of Business Administration in Karviná, Department of Mathematical methods in Economics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, tosenovskyfilip@opf.slu.cz.

[2] Silesian University in Opava/School of Business Administration in Karviná, Department of Mathematical methods in Economics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, CERGE-EI Affiliate Fellow, mielcova@opf.slu.cz.

The article is organized as follows: The next section is devoted to data description and basic unit root tests of the selected data. The third part describes the study of short-run and long-run dynamics; the fourth part discusses causality in data. A conclusion with list of references ends the paper.

2 Data description

To determine the mutual relationship between the Prague Stock Exchange and major stock exchanges, the data from the Dow Jones and S&P 500 indices from the U.S., the Japan's Nikkei 225 index, the Germany's DAX index and the Czech index PX-50 were used, namely their daily closing values for the period of 10/2012-4/2014 [17, 18]. The selected period covers more than 350 observations and does not include sudden changes that occured in the last recession in Europe in 2008-2012 [12]. For illustration, Figure 1 gives the series of the data.

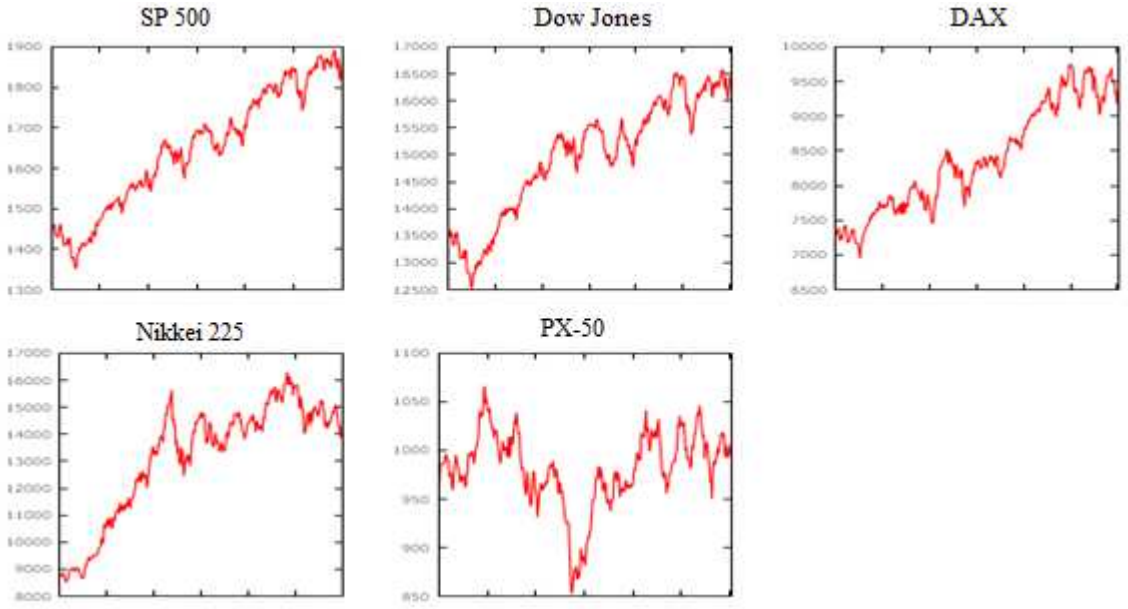


Figure 1 Data on Dow Jones, S&P 500, Nikkei 225, DAX and index PX-50. Source: Own calculations.

The main aim of the next analysis is to determine whether there exist a long-run as well as a short-run dependence between the data series. Since we intend to use either a vector autoregression model or a vector error correction model to investigate the long-run relationship among the time series, we first need to test a stationarity of the process for each variable, which means to use the Dickey-Fuller test for unit roots in the data. The Dickey-Fuller is used to determine if the variable ρ in the equation

$$y_t = \beta_0 + \rho.y_{t-1} + \varepsilon \tag{1}$$

is not statistically different from 1. Dickey and Fuller (1976) [4] rearranged equation (1) so that

$$y_t - y_{t-1} = \beta_0 + (\rho - 1).y_{t-1} + \varepsilon = \beta_0 + \beta_1.y_{t-1} + \varepsilon. \tag{2}$$

Then, ordinary least squares regression estimation of the first differences of the time series on the first lags of the series can show whether the estimated coefficient of the lagged variable is significant. If the coefficient is not statistically significant, the time series is expected to be stationary in the first differences, the hypothesis of a unit root cannot be rejected. Results of the estimation of equation (2) with t-ratios and p-values for each value are given in Table 1. Results indicate that the hypothesis of a unit root cannot be rejected on 1% level of significance; however, the variables of Nikkei 225 and PX-50 indices can be rejected on 10% and 5% level of significance, respectively. These two variables were re-examined using augmented Dickey-Fuller test of the form:

$$\Delta y_t = \beta_0 + \beta_1.y_{t-1} + \beta_2.\Delta y_{t-1} + \varepsilon \tag{3}$$

The null hypothesis (unit root) is based on test statistics equal to the t-ratio of the coefficient:

$$DF_t = \frac{\beta_1}{SE(\beta_1)} \tag{4}$$

the critical values are -3.44 and -2.87 for 1% and 5% level of significance, respectively.

	β_0	β_1 (p – value)	t-ratio
Dow Jones	80.52	-0.0048 (0.3555)	-0.9252
S&P 500	6.53	-0.0032 (0.4729)	-0.7186
Nikkei 225	158.21	-0.0108 (0.0512)	-1.952
DAX	55.10	-0.0059 (0.2997)	-1.039
PX-50	26.38	-0.0268 (0.0303)	-2.175

Table 1 Results of Dickey-Fuller tests estimation based on equation (2) Source: Own calculations.

Results of augmented Dickey-Fuller test for Nikkei 225 and PX-50 indices are given in Table 2; the hypothesis of unit root cannot be rejected for both variables.

	β_0	β_1	β_2	DF_t
Nikkei 225	162.21	-0.0110	-0.0813	-1.981
PX-50	29.13	-0.0296	0.1090	-2.396

Table 2 Results of augmented Dickey-Fuller tests based on equation (3) Source: Own calculations.

3 Long and short-run relationship

One of the objectives of the paper is to confirm or reject the general expectation related to what extent and with what delay the Prague Stock Exchange responds to developments in other financial markets. As the first step of the analysis, ordinary - least - squares linear model of the Prague Stock Exchange index PX-50 on all other indices was performed, *PX-50=f(Dow Jones, S&P 500, Nikkei 225, DAX)*. There were multicollinearity issues indicated in the model; variance increase factors of all the explanatory variables were higher than 10. Moreover, there is a high correlation between the explanatory variables – the smallest of the correlation coefficients reached a value of 0.85, all correlation coefficients were statistically significant at 5% level of significance. To avoid problems with multicollinearity, only pairwise comparisons of the PX-50 and all other selected stock-exchange indices were performed.

In order to test the existence of pairwise cointegrations between the PX-50 variable and all other variables, the Engle-Granger test was used [5]. The test is based on the test of a unit root in residuals of cointegrating regression. Thus, the test is divided into two steps: in the first step, ordinary regression estimation is performed on the variables; residuals of the models are stored to be used in the second step. As the second step, the Dickey-Fuller unit root test on residuals from step 1 is performed. Cointegration occurs if errors do not have a unit root, i.e. if the coefficient of lagged residual is statistically significant. The Dickey-Fuller unit root test is of the form:

$$\Delta res_t = \beta_0 + \beta_1 . res_{t-1} + \varepsilon \,. \tag{5}$$

Results of the cointegration regression and respective unit root tests for the stock exchange index variables are given in Tables 3 and 4.

Model	β_0	β_1	Residuals' variable
$PX\,50 = \beta_0 + \beta_1 . Dow\,Jones$	996.71	-0.0010	res_{DJ}
$PX\,50 = \beta_0 + \beta_1 . S\&P500$	951.05	0.0182	res_{SP}
$PX\,50 = \beta_0 + \beta_1 . Nikkei\,225$	1007.09	-0.0020	res_N
$PX\,50 = \beta_0 + \beta_1 . DAX$	877.45	0.0124	res_{DAX}

Table 3 Regression results for ordinary least squares models. Source: Own calculations.

	β_0	β_1 ($p-$value)	DF_t
Dow Jones	0.1267	-0.0270 (0.0297)	-2.183
S&P 500	0.0970	-0.0266 (0.0307)	-2.170
Nikkei 225	0.1493	-0.0279 (0.0264)	-2.230
DAX	0.0460	-0.0265 (0.0307)	-2.170

Table 4 Results of Dickey-Fuller tests based on equation (5). Source: Own calculations.

All coefficients β_1 from Table 4 are statistically significant at 5% level of significance; also, comparison of DF_t values with critical values of Dickey-Fuller test indicates that the null hypothesis of unit root can be rejected on 5% level of significance. Thus, results indicate that all pairs of studied variables have a long-run equilibrium relationship.

With a cointegration present in the time series, we can estimate the error correction model which focuses on the short-term dynamics. The number of lags in the model was set to be 1, as was indicated by the Hannan-Quinn criterion and the Schwartz-Bayes criterion; Akaike information criterion suggested 2 - lags estimation. The respective error correction model is

$$\Delta PX50 = \beta_0 + \beta_1.\Delta PX50_{t-1} + \beta_2.\Delta INDEX_{t-1} + \beta_3.res_{t-1} + \varepsilon\,. \tag{6}$$

Results of the estimation for all studied relations are given in Table 5; all estimated coefficients of the residual variables β_3 are statistically significant at 5% level of significance for all the four models.

	β_0	β_1 ($p-$value)	β_2 ($p-$value)	β_3 ($p-$value)	$\frac{1}{\lvert\beta_2\rvert}$
Dow Jones	33.04	0.0666 (0.2280)	0.0121 (0.0100)	-0.0303 (0.0127)	33.0
S&P 500	29.53	0.0610 (0.2663)	0.1181 (0.0027)	-0.0293 (0.0163)	34.1
Nikkei 225	29.38	0.1225 (0.0230)	-0.0036 (0.0911)	-0.0260 (0.0201)	38.5
DAX	28.95	0.1066 (0.0654)	0.0007 (0.9127)	-0.0299 (0.0172)	33.4

Table 5 Estimated coefficients of error-correction models based on equation (5). Source: Own calculations.

Error-correction term (estimated coefficient β_3) in all four estimations is negative and smaller than 1, which supports the appropriateness of the error-correction model. Between all studied pairs of indices, there is a long-term relationship, a positive error term from that long-term relationship will be counteracted in a subsequent period by approximately 2.60 % – 3.03% (coefficient β_3). The positive shock in a long-run relationship (that means a positive residual from the long-term relationship) will be corrected in the next period and will be completely absorbed after approximately 33-38 periods (last column of Table 5); in this case, approximately after about seven weeks. For illustration, the impulse – response functions plots of the change in PX-50 after a unit change of the Dow-Jones index and S&P 500 index, respectively, are given in Figure 2.

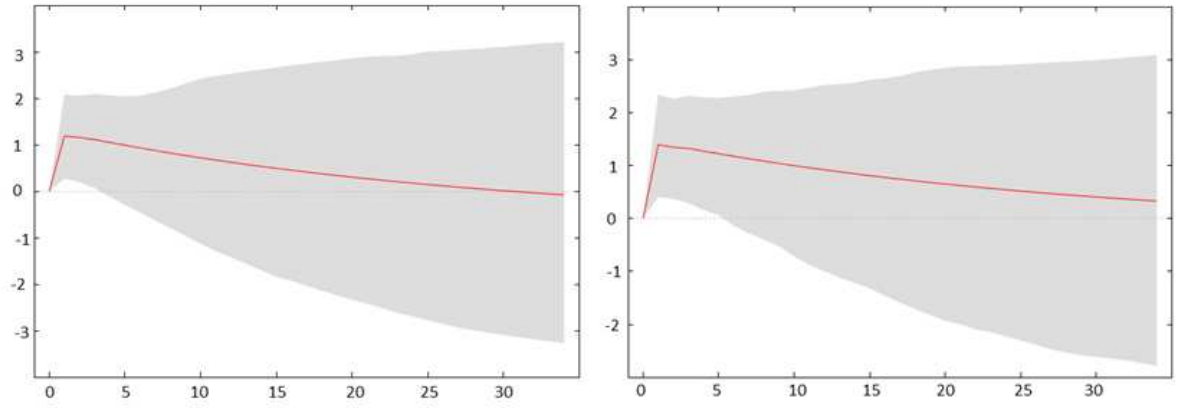


Figure 2 Impulse-response functions of the change in PX-50 after a unit change of the Dow-Jones index and S&P 500 index, respectively, with 95% bootstrap confidence interval. Source: Own calculations.

4 Causality

The short-run causality was studied by the Granger causality test. In general, the Granger causality [16] test is constructed such that it reveals time changes more than causality; the meaning of the expression "Granger cause" is equivalent to the expression "precede" more than "cause". The co-integration test – the Granger causality test is based on comparison of unrestricted and restricted VAR models, estimated without additional explanatory variable:

$$F = \frac{(RSS_{\text{restricted}} - RSS_{\text{unrestricted}})/M}{RSS_{\text{unrestricted}}/(N-K-1)} \approx F(M, N-K-1), \tag{7}$$

where K is the number of unrestricted model slope coefficients, M equals the number of slope coefficients eliminated in the restricted equation, N is the number of observations and RSS denotes the sum of squared residuals. Results for pair-wise comparisons (2 lags or 1 lag) of changes in index developments are given in Table 6 with values of F-test and p-values.

Statement	Lags	F test	p-value	Result for $\alpha = 0.05$
"$\Delta DowJones$ Granger causes $\Delta PX\,50$"	2	3.6669	0.0265	Yes
"$\Delta S\&P500$ Granger causes $\Delta PX\,50$"	2	4.7538	0.0092	Yes
"$\Delta Nikkei225$ Granger causes $\Delta PX\,50$"	2	1.3714	0.2551	No
"ΔDAX Granger causes $\Delta PX\,50$"	2	0.5911	0.5543	No
"$\Delta Nikkei225$ Granger causes $\Delta PX\,50$"	1	2.9388	0.0874	No
"ΔDAX Granger causes $\Delta PX\,50$"	1	0.0019	0.9649	No

Table 6 Results of Granger causality tests. Last column indicates results $\alpha = 0.05$. Source: Own calculations.

The results show only two Granger causality relations, development in Dow Jones index and S&P 500 index precedes development in PX-50 index. The respective impulse – response functions plots of the change in PX-50 after a unit change of Dow Jones and S&P 500 are given in Figure 3.

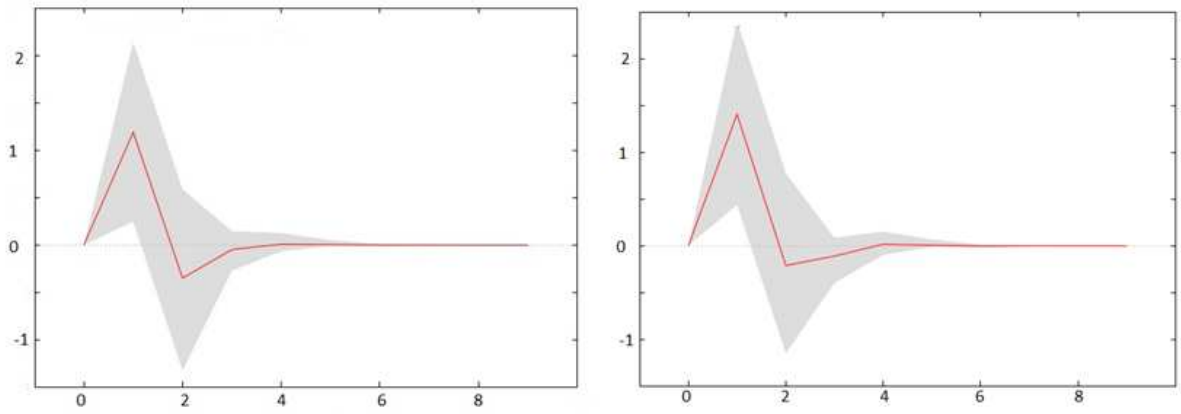


Figure 3 Impulse-response functions of the change in PX-50 after a unit change in Dow Jones and S&P 500 index, respectively, with 95% bootstrap confidence interval. Source: Own calculations.

5 Conclusions

The aim of this paper was to discuss a general expectation related to what extent and with what delay the Prague Stock Exchange, as a minor stock market, responds to developments in other financial markets. In order to reveal the mutual relationship between the Prague Stock Exchange and major stock exchanges, we examined daily closing values of the Dow Jones, S&P 500, Nikkei 225 and DAX indices and compared them with the Czech index PX-50. The data series are I(1) stationary, we made pairwise comparisons of the index PX-50 with each of the aforementioned indices. Results of Engle-Granger tests indicate that all pairs of studied variables

have a long-run equilibrium relationship. Therefore, with present cointegration in time series, we could expect short-term dynamics, described by an error-correction model. The respective error-correction models indicated that error-correction terms are about 2.60 %-3.03%, therefore the positive shock between all pairs of variables would be completely absorbed after about seven weeks.

We tried to find a cointegrating relationship, tested by the Granger causality test. The results partially support the expected influence of major capital markets on the minor market; development in the Dow Jones and S&P 500 indices precedes that of PX-50 index, however this relation is not true for Nikkei 225 and DAX indices.

Acknowledgements

This paper was supported by the project "Innovation of Educational Programs at Silesian University, School of Business Administration in Karviná" n. CZ.1.07/2.2.00/28.0017.

References

[1] Alexander, C.: Optimal Hedging Using Cointegration. In: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **357**(1999), 2039-2058.
[2] Batini, N. and. Greenslade, J. V.: Measuring the UK Short-Run NAIRU. *Oxford Economic Papers* **58** (2006), 28-49.
[3] Chen N.-F., Roll R. and Soss S.A.: Economic Forces and the Stock Market. *The Journal of Business* **59** (1986), 383-403.
[4] Dickey D. A. and Fuller W. A.: Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association* **74** (1979), 427-431.
[5] Engle, R. F. and Granger, C. W. J.: Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* **55** (1987), 251-276.
[6] Faff, R. W. and McKenzie, M. D.: The Impact of Stock Index Futures Trading on Daily Returns Seasonality: A Multicountry Study. *The Journal of Business* **75** (2002), 95-126.
[7] Goyenko, R. Y. and Ukhov, A. D.: Stock and Bond Market Liquidity: A Long-Run Empirical Analysis. *The Journal of Financial and Quantitative Analysis* **44** (2009), 189-212.
[8] Harris, F. H. de B., Mcinish, T. H., Shoesmith G. L., and Wood R. A.: Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets. *The Journal of Financial and Quantitative Analysis* **30** (1995), 563-579.
[9] Hiemstra, C. and Jones, J. D.: Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation. *The Journal of Finance* **49** (1994), 1639-1664.
[10] Kim, Y.: Purchasing Power Parity in the Long Run: A Cointegration Approach. *Journal of Money, Credit and Banking* **22** (1990), 491-503.
[11] King, M., Sentana, E. and Wadhwani, S.: Volatility and Links between National Stock Markets. *Econometrica* 62(1994), 901-933.
[12] Mazurek, J. and Mielcova E.: The Evaluation of Economic Recession Magnitude: Introduction and Application. *Prague Economic Papers* **2** (2013), 182-205.
[13] Miller, S. M.: Monetary Dynamics: An Application of Cointegration and Error-Correction Modeling. *Journal of Money, Credit and Banking* **23** (1991), 139-154.
[14] Psaradakis, Z., Sola, M. and Spagnolo, F.: On Markov Error-Correction Models, with an Application to Stock Prices and Dividends. *Journal of Applied Econometrics* **19**(2004), 69-88.
[15] Stoll, H. R. and Whaley, R. E.: The Dynamics of Stock Index and Stock Index Futures Returns. *The Journal of Financial and Quantitative Analysis* **25**(1990), 441-468.
[16] Wei, W. W. S.: *Time Series Analysis. Univariate and Multivariate Methods*, Addison-Wesley Publishing Co., 75 Arlington Street Suite 300, Boston MA 02116 USA, 1994.
[17] WWW: *Yahoo! Finance* [online]. [cit. 2014-05-05]. Available at WWW: http://finance.yahoo.com/stock-center/.
[18] WWW: *Kurzy.cz* [online]. [cit. 2014-05-05]. Available at WWW: http://akcie-cz.kurzy.cz/burza/index-px/historie/.

A simulated heterogeneous agent double auction stock market

Quang Van Tran[1], Jan Kodera[2]

Abstract. In this experiment, we try to simulate a financial market with three types of traders: fundamentalists, chartists and noisy ones. The market uses a continuous double auction to match orders and enable the trading of one stock. The orders are of two types: the limit ones are submitted by patient traders and the less patient ones submit the so called market orders. While limit orders can only guarantee the price, but not execution, market orders can always secure the execution at the best available price. In the model, we allow the fundamental value of a stock evolve as a jump process as a result of occasional news arrivals. We assume the fundamentalists may know the underlying value of the stock while the noisy traders base their decision on some noisy signal and the chartists make their decision according to the trend. Under this market set-up the trade occurs and a stock price series can be generated. We study the statistical properties of this series and compare them to the ones of the real price series. We also investigate how the size of the market parameters can affect these properties.

Keywords: Simulation, Heterogeneous agent, Double auction, Financial market

JEL classification: C15
AMS classification: 68U20

1 Introduction

The simulation of artificial financial market is a useful tool to analyze how financial markets work and it has become an extensive subject of interest of researchers [3, 4, 5]. Though simulation cannot shed much light on how agents disseminate risk, it definitely allow us to look into their interactions when trading. In this study, through simulations we investigate the behavior of a financial market with three types of agents: fundamentalists, chartists and noisy traders. Each type of agents behaves in an expected way typical for them and trades each with other on a profit motive. The trades are set with a double auction mechanism in which the trader's order is matched with the best existing bid/ ask prices. Further, based on his own valuation of the stock, a trader can also submit his own offer and if meets all requisites, it will be included in the order book. When offers are executed, the stock prices are settled and over the course of time this market setting can generate a series of stock prices which is an important source of information for analyzing fundamental features of market behavior. To simulate a financial market, a number of parameters have to be set. In this paper we try to examine how the values of these parameters can affect the artificial financial market measured by the characteristics of prices series.

2 Double auction on a stock unit

Auctions have a long history as a mean to buy or sell an object. It is a process, in which an item is offered for buy or sale and bidder can submit their bids. These competitive bids are recorded and the item is allocated to the bidder with the highest bid in a sale auction and to the bidder with the lowest ones in a buy auction. In comparison with a one-on-one haggling procedure, the advantage of an auction lies in

[1]University of Economics, Prague,Department of Statistics and Probability, Winston Churchill Sq. 4, Prague, Czech Republic, tran@vse.cz
[2]University of Economics, Prague,Department of Statistics and Probability, Winston Churchill Sq. 4, Prague, Czech Republic, kodera@se.cz

the fact that the seller or buyer is supposed to have a bunch of applicants competing for the item given that they are unable to collude. This competition may push the price of the object to the highest level in a sale auction or to the lowest level in a buy auction. Compared to a posted price sale or purchase, an auction can give some additional information to the owner (buyer) on how bidders value the object to alleviate his or her uncertainty about his or her own valuation of the object of interest. Other advantages of an auction mechanism are its universality and anonymity, e.i. auctions can be used to buy or sell any object and bidders identity plays no role in the item allocation. For all these reasons auctions are used as an allocating mechanism in financial markets (for more information on auction, see [6]).

There are several ways how to organize an auction. It can be the so called second price auction when in a sale auction the price of a stock bidders submit is what each bidder think that it should be. And the highest bid wins and the winner pays the second highest price. Organizing the auction this way, the seller can obtain information on how bidder value the stock. The other way to organize an auction is the first price auction when the highest bid wins and the bidder pays his bid. In this case each bidder will submit a price whose value must be lower than the value of a stock he wants to buy. For financial markets the latter seems to be the appropriate case as the stock is allocated to the highest bid price.

An auction can be either a sealed bid auction or an out cry auction. In both cases, as far as the valuation of a stock is concerned, the value can be absolutely private when a bidder knows it with certainty and therefore his bid is not affected by the valuations of other bidders. But it may happen that a bidder can possess only partial information on the value of the stock. Then his valuation might be affected by the valuation of other bidders. In this case we have interdependent valuation. As the exact value of a stock is unknown, the stock value interdependence may make a bidder to update his valuation during the course of an auction. As a result, he may eventually win the auction, but at the same time he also overestimate the value of a stock which constitute the so called the winner curse.

In our setting a financial market consists of many buyers and sellers who are endowed either with cash or stock. At any time they can make offers either to buy or sell or to accept other traders' offers on one stock. As a result, the stock market works as a double open out cry auction with interdependent valuation, one for sale of one stock and the other for buy one stock. The stock market continuously tries to match these offers in the market to cause a trade happen between a buyer and a seller. A buying offer is called a bid and a selling offer is called an ask and these offers are stored in an order book for a certain period. During a trading time, a trading agent may post any ask order and accept any bid order at any time and buyers can post any bid order and accept any ask order at any time. If one buyer's bid matches a seller's ask on price, an exchange will occur and the stock exchanged will no longer be available at this price anymore.

Together all bids and asks form a bid order book and an ask order book. After each transaction the order is cleared from the order book and the second best order becomes the actual one. As agents can only see the best bid and the best ask at any given time in the trading period, for a new bid (ask) to be the best bid (ask), a buyer (seller) must submit a bid (ask) that is higher (lower) than the best bid (ask). This rule is called bid/ask improvement rule. Since Agents can observe the best bid and ask, throughout the trading time the difference between the highest bid and the lowest ask, which is called the spread, is known to all agents in the market. Therefore, they always know how to make their bids become the best ones. Further, the history of trades in previous trading periods is available to all market participants as the historical price series. Also at the end of each trading day the order book is cleared and a new day comes with a new order book.

3 The model with heterogeneous agents

As we have already mentioned the market is consisted of three types of agents. The first one are the fundamentalists. They are supposed to know the fundamental value of the stock and at the beginning of the experiment this value is assumed to be V_0. During a trading day, the fundamental value of the stock fluctuates around V_0 and can jump to either direction with a jump of constant size with a non-negative probability, which can be formalized as follows

$$V = V_0 + \xi_1 + p_j * \xi_2, \tag{1}$$

where $\xi_1 \sim N(0, \sigma_1^2)$, $\xi_2 \sim N(0, \sigma_2^2)$, $\sigma_2 > \sigma_1$, and p_j is the probability that jump will occur. As the fundamentalists know the underlying value of the stock, they can either submit a bid to buy the stock if

they observe the existing ask price lower then its value of the stock and vice versa, that is if they can sell the stock (ask offer) they have if seeing the bid price higher than its value. If the existing bid/ask price does not meet the conditions mentioned above, the fundamentalists by themselves can make a bid/ask offer in such way that the bid price must be lower than it underlying value and the ask price has to be higher than its fundamental value. Further, for a new trading day the fundamental value of the stock is set as the last value from previous trading day.

In accordance with common belief on chartists, the price of the stock is determined in the following way. According to the chartists the stock price evolves over time in trend, they observe the price of the stock recorded in the price series and the price for the next trading period will follow the trend from the previous periods and in this case it always is the moving average of the last few period prices, that is

$$P_t = \frac{1}{k}\sum_{i=1}^{k} P_{t-i} + \xi_3, \tag{2}$$

where k is the number of period taken into account for calculating the trend and ξ_3 is some arbitrary noise specific to each chartist trader. Similar to the fundamentalists, a chartists would submit a bid if he sees the ask price lower than his own valuation of the stock. On the contrary, if he observes the bid price higher than his own valuation of the stock unit, he would make an ask offer to sell his stock. If there is no such offer satisfying these conditions, he can submit his own bid/ask offer accordingly in a similar fashion as the fundamentalists.

The last type of traders in the experiment are noisy traders. A noisy trader neither knows the fundamental value of the stock nor uses the trend to evaluate the stock value. In our experiment the noisiness of his price signal is introduced by the assumption that the stock price signal of a noisy trader is calculated as the initial underlying value of each trading day mixed up with some noise, that is

$$S_t = V_t + \xi_4, \tag{3}$$

where $\xi_4 \sim N(0, \sigma_4^2)$ and $\sigma_4 > \sigma_1$. Again like a fundamentalist and a chartists, a noisy trader would submit a bid if he see the ask price lower than his own signal on the stock price. On the contrary, if he observes the bid price higher than his own signal of the stock unit, he would make an ask offer to sell his stock. If he does not observe any quote meeting these conditions, he can submit his own bid/ask offer accordingly in a similar fashion as both previous types of traders. In all three cases noise is added to express the fact that the stock price is specific to each trader. On the other hand, as the noise is gaussian, the price often does not deviate from its average value.

4 The simulation and its result

In order to run the simulation, a set of parameters needs to bet set. First we assume that a trading day lasts 8 hours which is 8x60x60 seconds. A series of 300 trading days is simulated. As we do not consider that each trader would participate in submitting an offer in every second, this activity will occur with probability p_1 in a second. Then for each trading second of a day a number is drawn from a uniform distribution $U(0, 1)$. If the number is less than p_1, the trading activity will happen at this time interval. We also assume that an offer once submitted and if no exchange happens during its lifetime or no better offer occurs can only last 10 minutes. After that it is automatically deleted.

Further, a trader can be a fundamentalist with probability p_f, a chartist with probability p_c and finally a noisy trader with probability p_n. It is obvious that p_f, p_c and p_n must be positive and $p_f + p_c + p_n = 1$. These probabilities also are the share of each portion of traders of some type in the whole population of traders. As we assume the existence of jump in the fundamental value of the stock, the probability of the occurrence of jump is p_j. It is a small positive number as the jump is supposed to happen very rarely. Since the stock price is assumed to be specific to each trader, the magnitude of a price change is affected by the size of variances of corresponding noise ξ_1, ξ_2, ξ_3, and ξ_4.

In figures 1 to 4 the development of the stock price, its returns, the histogram of it returns, and a part of the intraday development of prices are respectively shown for this set of parameter values: $V_0 = 1$ [1], $p_1 = 0.9$, $p_f = 0.4$, $p_f = 0.3$ and $p_f = 0.3$. The magnitudes of noise variance are $\xi_1 = 0.005$, $\xi_2 = 0.05$,

[1] The value of V_0 is nothing but a normalized value of any stock price.

$\xi_3 = 0.05$, and $\xi_4 = 0.01$. The outputs of the simulation exercise seem to show similar characteristics of an actual stock in reality. The daily stock price seems to be a random walk. The corresponding returns are fluctuating around 0 with a few "extreme" values. The histogram of returns resembles the one of normal distribution. Intraday returns in Figure 4 are lower than daily returns. We have experimented with parameter p_1 and the results show that the value of this parameter seems to not play very important role in shaping the development of the stock price. At the moment we have found that it can only affect the number of trades executed in one trading day. At this stage our focus is put on how the share of each type of traders can affect the stock price dynamics.

In figures 5 to 7 the stock price dynamics is displayed for the case when the share of fundamentalists is increased to 90 % and the shares of chartists and noisy traders are minimal, that is $p_c = p_n = 0.05$. The values of other parameters are kept unchanged. In this case, traders are predominantly of fundamentalists. The results of simulation show that increasing the share of fundamentalists does not lead to reduction of trading. On the other hand, though random walk nature of the price development remains, the magnitude of stock returns is substantially reduced. Visually, no extreme values are observed. The shape of the histogram seems to lack the typical leptokurtic attribute of returns of financial assets.

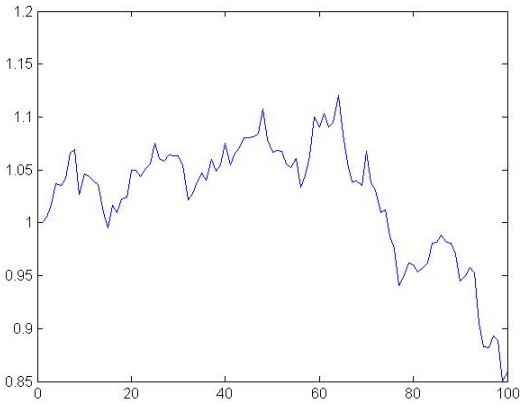


Figure 1 The daily stock price

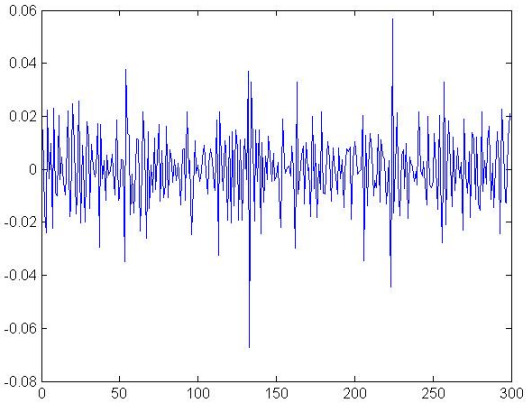


Figure 2 The daily stock returns

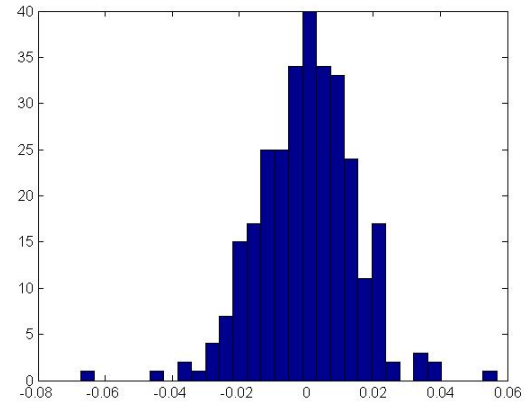


Figure 3 The daily stock returns histogram

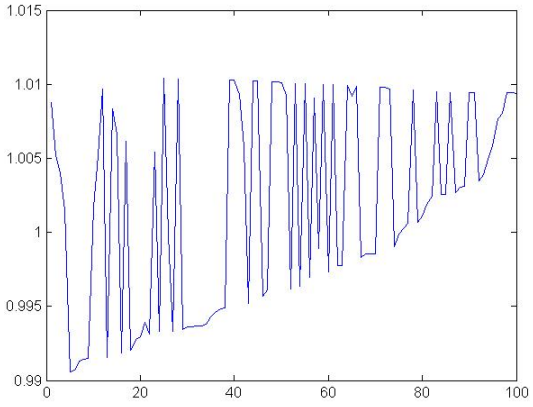


Figure 4 The intraday stock prices development

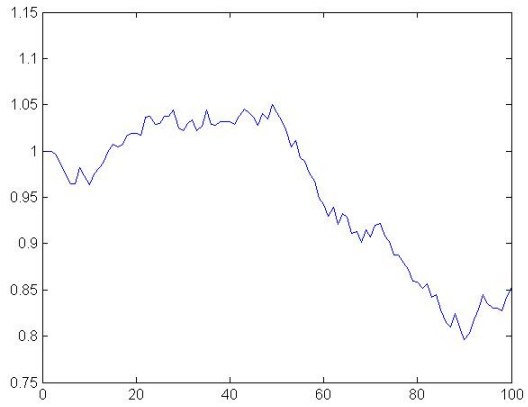


Figure 5 The daily stock prices development II

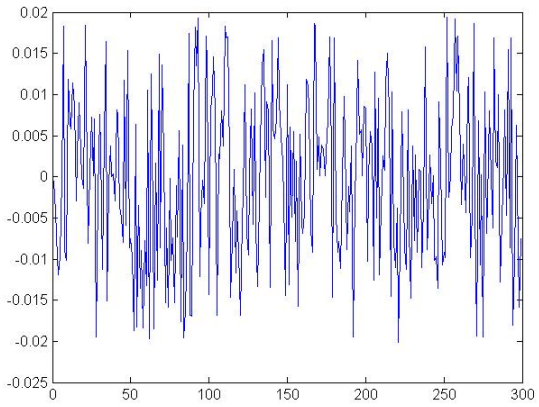


Figure 6 The intraday stock prices development

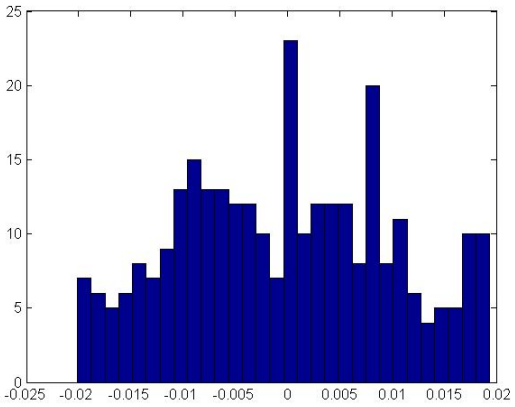


Figure 7 The histogram of daily stock returns

5 Conclusion

In this experiment we have tried to simulate a behavior of price of a stock in a financial market with three types of traders: fundamentalists, chartists and noisy traders. These traders exchange a stock for cash with a profit motive. The exchange operates through a double auction mechanism. Each trader can submit an offer appropriate to the situation when the existing bid/ask price is profitable for him to make such an offer. On the other hand if no such price exists, he can make his own bid/ask offer which could potentially generate him some profit with respect to his own valuation of the stock. Under this setting of the market we find that this trading mechanism can generate a price dynamics which is similar to the common dynamics of a stock price series in reality. We have also found that under this arrangement of the market, the share of each type of traders can substantially affect the price dynamics, namely the range of the returns generated by trading. Keeping all other parameters unchanged, the higher the share of fundamentalists is, the lower the returns are. In our opinion, this study is only the first step of how simulations can be used to examine behavior of financial markets and it can be extended to many directions.

Acknowledgements

The authors are grateful for the financial support of Czech Science Foundation No. P402/12/G097.

References

[1] De Long B. J., Shleifer A., Summers L. H., Waldmann, R. J.: Noise trader Risk in financial market. *Journal of political economy* **98** (1990), 703 –738.

[2] Boswijk P. H., Cars H. Hommes C. H., Manzan S.: Behavioral heterogeneity in stock prices. *Journal of Economic Dynamic & Control* **31** (2007), 1938 – 1970.

[3] Chiarella, C., Iori, G.: A Simulation Analysis of the Microstructure of Double Auction Markets. *Quantitative Finance* **2**, 5(2002), 346–353.

[4] Dezsi, D., Maries, I., Balu, F. O.: Heterogeneous Agents' Interactions in a Double Auction Virtual Stock Market. *Computational Collective Intelligence.* Technologies and Applications Lecture Notes in Computer Science **8083** (2013), 205–214.

[5] Guo, T. X.: An agent-based simulation of double-auction markets. *Masters' Thesis*, University of Toronto, 2005.

[6] Krishna, V.: *Auction Theory*, 2nd edition, Academic Press, 2009.

[7] LeBaron, B.: Agent-based computational finance: Suggested readings and early research. *J. Econom. Dynam. Control* **24** (2000), 679–702.

[8] Madhavan, A.: Market Microstructure: A Survey. *Journal of Financial Markets* **3** (2000), 205–258.

[9] O'Hara, M.: *Market Microstructure Theory.* Blackwell, Malden, MA, 1995.

[10] Peters, E. E.: *Fractal Market Analysis, Applying Chaos Theory to Investment and Economics*, Springer, NY, 1994.

Poverty Analysis and Measuring Income Inequality in Czech Republic

Kamila Turečková[1], Eva Kotlánová[2]

Abstract. Every human society which is formed in a framework of the specific economic structure is characterised by different poverty rate, social cohesion and income inequality. Economic well-being and standards of living which could be measured for example by GDP per capita or through average wages, does not predicate real spread of incomes in the society. This allocation is mostly considerably unequal. Income inequality creates a space for social and economic discrimination. One of the best known and used measures of income inequality is Gini coefficient and its graphical representation – Lorenz curve. It could be supplemented with Robin Hood Index which is used as a method of comparison of income inequality.
The aim of this paper is to use Gini coefficient, Lorenz curve and Robin Hood Index to map the changes in income inequality of Czech inhabitants in the period of 2004 – 2011 with respect to economic and social changes, which could affect the incomes. The analysis of incomes allocation of Czech households and its order will be made in deciles based on empirical data of Czech Statistical Office.

Keywords: Gini coefficient, income inequality, Lorenz curve, Robin Hood Index.

JEL Classification: C13, D31, I32,
AMS Classification: 62P20

1 Introduction

There are many possibilities how to look on or measure standards of living in selected countries. One of the best known is GDP per capita. Despite the fact that this indicator could reach relatively large value, it does not predicate about differences of incomes in society. Another indicator we could hear about very often is average wage. Not even its amount is guarantee of economic well-being. It is usual that over 50% of working population of the country could not reach this amount. One of the best known and used measures of income inequality is Gini coefficient and its graphical representation – Lorenz curve. It could be supplemented by Robin Hood Index which is used as a method of comparison of income inequality.

The aim of this paper is to use Gini coefficient, Lorenz curve and Robin Hood Index to map the changes in income inequality of Czech inhabitants in the period of 2004 – 2011 with respect to economic and social changes which could affect the incomes. Why is measuring of income inequality so important? Some studies found the relationship between income inequality and economic growth [4], or income inequality and society's level of mortality [3] and income inequality could impact many others economic or social indicators.

2 Measurement of Income Inequality

Income inequality is a natural part of every human society. It could reflect a measure of poverty and redistribution of income. Among methods how to measure income inequality are mostly included: Lorenz curve, Gini coefficient, Atkinson index, Robin Hood index and Variation coefficient [1], [8]. In this paper we will use Lorenz curve, Gini coefficient and Robin Hood index, so we describe them closer. For more information about other indexes see [1], [5], [8] or [9].

2.1 Lorenz curve

Lorenz curve is probably the most widely used method of graphical expression of income inequality. It is constructed in a box diagram with two scales, each 0 – 100%. Every point on the Lorenz curve indicates share of income-receiving unit (mostly households) and their share of the total income. Equality of income distribution in

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Economics, Univerzitní nám. 1934/3, Karviná 733 40, tureckova@opf.slu.cz.
[2] Silesian University in Opava, School of Business Administration in Karvina, Department of Economics, Univerzitní nám. 1934/3, Karviná 733 40, kotlanova@opf.slu.cz.

society is found when Lorenz curve is identical with a straight line with angle of 45°, it means that every income-receiving unit receives its proportional share of the total income, see Figure 1. Let's define inequality as any deviation from equality. Thus, if any person received more than his proportionate share of the aggregate income, the distribution would be unequal [8]. In a case of an absolute inequality of income distribution the Lorenz curve will be identical with graphs axis. It means that all incomes will receive just the only one person.

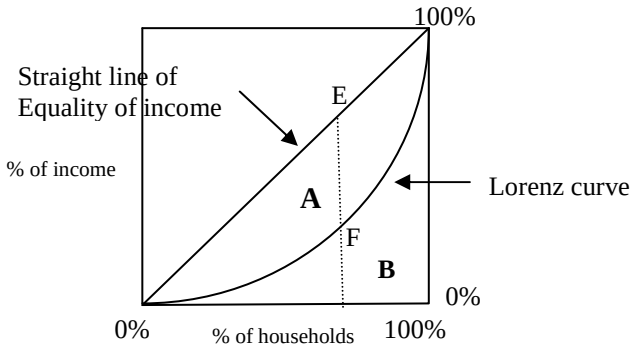


Figure 1 Lorenz curve

It is possible to formulate Lorenz curve as a distribution function F(x) as an equation:

$$L(F)=\frac{\int_{-\infty}^{x(F)} x\cdot f(x)\cdot dx}{\int_{-\infty}^{\infty} x\cdot f(x)\cdot dx}=\frac{\int_{0}^{F} x(F')\cdot dF'}{\int_{0}^{1} x(F')\cdot dF'}, \tag{1}$$

where x(F) is inverse function to the distribution function F(x).

Through graphical representation of Lorenz curve it is possible to illustrate another two types of measurement of income inequality – Gini coefficient and Robin Hood index.

2.2 Gini coefficient

The most widely used method of measure of concentration in income inequality area is Gini coefficient which is derived from Lorenz curve and it is a numeric representation of divergence of Lorenz curve from Straight line of equality. The Gini ratio of concentration is calculated by dividing A area and a sum of A and B area (see Figure 1).

$$G=\frac{A}{A+B} \tag{2}$$

The value of Gini coefficient can range from 0 to 1, which means that the 0 represents a total equality of incomes distribution and value 1 total income inequality, when all income in economy receives the only one person. In statistics we usually find Gini coefficient expressed as a percentage.

The main advantage of using this coefficient is its easy interpretation and the fact that it measures income inequality regardless of a population in the country and regardless of a size of economy. It could be used to compare of income distribution among different groups of population as well as single regions in one country [9].

Objections to the crudity and ambiguity of this measure arise from the fact that every Lorenz curve has just the only one Gini coefficient but two societies with the same Gini coefficient could have different Lorenz curve. In view of this fact it is recommended to supplement results of Gini coefficient with conclusions of another analyses or indexes.

2.3 Robin Hood Index

It is known as Hoover or Schutz index. Graphically it is derived from Lorenz curve. Look at the Figure 1, it is a distance between points E and F, which is probably the longest vertical distance between Lorenz curve and the straight line of equality of income. The value of Robin Hood Index indicates how large should be the transfer of

income leading to equality of income, i. e. which part of income in society is necessary to redistribute from richer households to the poorer to achieve equality in incomes distribution (the value of the index approaches zero).

Procedure of computation is consequential: It is necessary to classify households according to value of their income from the richest to the poorest and then to divide them into 10 groups of the same size. After this the sum of their percentage shares of total income is calculated, but just for the groups of households which share is more then 10%. Finally it must be subtracted n-multiple of 10%, where n is a number of households which income is higher than 10%.

3 Methodology and data

From a methodological perspective, the work is based on graphical analyses of Lorenz curve and calculations of Gini coefficients and Robin Hood Index for Czech Republic. All these measures of income inequality were described above. The software used was MS Excel.

For construction of Lorenz curve and calculation of indexes were used households net money incomes which were calculated as:

$$E = A - B - C + D \tag{3}$$

Where ….. E is Net money income
….. A is Households gross money income
….. B is Contributions to mandatory social security schemes
….. C is Income tax
….. D is Tax bonus

Much of the data was taken from the Czech Statistical Office database [2]; an additional source was the Ministry of Labour and Social Affairs [7] from which the data of the minimum wage was taken and the data of income taxes was taken from web portal Ministry of Finance [6]. All calculations and graphical analyses are authors own. The covered period includes years 2004-2011 (credible data is not available for a longer period).

4 Empirical Analysis

Empirical analyses were made on the basis of the net money income of Czech households's data [2]. In Table 1 there are data for construction of Lorenz curves for each year.

households	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
2004	3.71%	9.37%	16.24%	23.94%	32.28%	41.36%	51.39%	62.87%	76.75%	100.00%
2005	3.87%	9.61%	16.55%	24.28%	32.63%	41.66%	51.60%	62.99%	76.75%	100.00%
2006	3.89%	9.72%	16.69%	24.42%	32.77%	41.86%	51.90%	63.33%	77.23%	100.00%
2007	4.07%	10.05%	17.05%	24.77%	33.11%	42.19%	52.20%	63.62%	77.41%	100.00%
2008	4.11%	10.15%	17.14%	24.77%	33.06%	42.06%	52.00%	63.30%	76.85%	100.00%
2009	4.02%	10.00%	17.01%	24.70%	33.08%	42.20%	52.26%	63.67%	77.36%	100.00%
2010	3.90%	9.84%	16.83%	24.51%	32.90%	42.03%	52.16%	63.72%	77.62%	100.00%
2011	3.91%	9.86%	16.90%	24.69%	33.17%	42.35%	52.49%	64.01%	77.84%	100.00%
straight line	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

Table 1 Households net money income (%) - cumulated

Because of the fact that the differences among Lorenz curves (2004 -2011) were not marked enough, Figure 2 illustrates Lorenz curves just for years 2004 and 2011, where the difference is visible. Lorenz curve 2011 is much closer to straight line than Lorenz curve 2004, which means, that came to an equalization of income inequality.

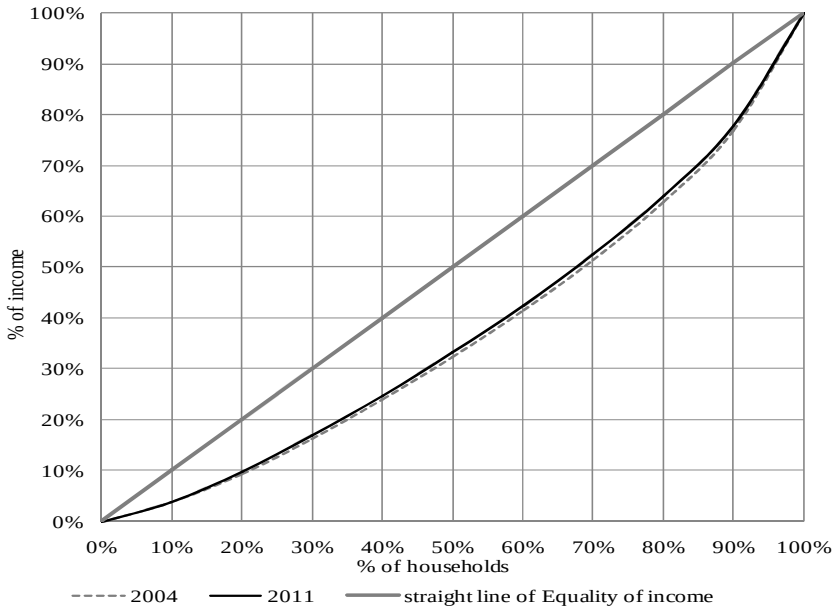


Figure 2 Lorenz curves for Czech Republic in 2004 and 2011

Subsequently on the basis of the same data, which were adjusted for computing of Gini coefficient and Robin Hood Index, these were calculated for the same period. As you can see in Table 2 results of Gini coefficient and Robin Hood Index correspond to Lorenz curves in Figure 2. The value of indexes decreased progressively excepting years 2008 and 2010.

Year	Gini coefficient	Robin Hood Index (%)
2004	0.2642	18.64
2005	0.2601	18.40
2006	0.2564	18.14
2007	0.2511	17.81
2008	0.2531	18.00
2009	0.2514	17.80
2010	0.2530	17.97
2011	0.2496	17.65

Table 2 Results of Gini coefficient and Robin Hood Index (2004 – 2011)

What were the reasons of increase of these indexes in 2008 and 2010? In 2008 the methodology of computation of the average wage and living wage was changed and in taxes area there the rate of natural person income tax changed too. You can see all these changes including progress of minimal wage in Table 3.

Year	Monthly average gross wage (CZK)	Basic rate of natural person income	Minimal wage (CZK)
2004	17,466	15%	6700
2005	18,344	15%	7185
2006	19,546	12%	7955
2007	20,957	12%	8000
2008	22,952	15%	8000
2009	23,344	15%	8000
2010	23,864	15%	8000
2011	24,455	15%	8000

Table 3 Selected indicator influencing income inequality in Czech Republic (2004 – 2011)

In years 2004 – 2007 the system of progressive taxation was applying in the Czech Republic, when the tax base was amount of gross wage decreased of contributions to mandatory social security schemes. The change in 2008 consisted of implementation of system of flat tax rate (15%) and the tax base has been changed too. It has been gross wage which is not decreased of contributions to mandatory social security schemes paid by employee and is increased by contributions to mandatory social security schemes paid by employer. So in reality people pay more than before because of higher tax base. Very important factor which is not involved in this analysis is a bonification. In 2010 consequences of economic crises took effect on living standard of people too. The rate of unemployment increased, so incomes of households decreased.

5 Conclusions

There exist many ways how to measure or describe income inequality in our society. In this paper we have paid attention to the most known of them – Lorenz curve, Gini coefficient and Robin Hood Index. They were used to map the changes in income inequality of Czech inhabitants in the period of 2004 – 2011. In selected period all indicators shows that trend leads to improvement in redistribution of incomes and income inequality in 2011 is better than in 2004. There are many factors which affect income inequality as minimal wage, amount of living wage or rate of taxation and redistribution. A special factor is an impact of economic cycle. Income inequality is falling away when economic is in recession and conversely.

Despite our results, data of many institutions looking into poverty show that more than one fifth of Czech inhabitants live on the border of poverty. So this small analysis is seen as an input for further research to be carried out by the authors. Next step is to analyse components, which could influence income inequality and poverty in Czech Republic, through regression analysis.

Acknowledgements

This paper was supported by the project OPVK Innovation of study programs at the Silesian University, School of Business Faculty in Karviná, Project registration number CZ.1.07/2.2.00/28.0017.

References

[1] Atkinson, A. B.: On the Measurment of Inequality. *Journal of Economic Theory* **2** (1970), 244 – 263.
[2] Czech Statistical Office: *Statistics on Income and Living Conditions*. 2014. [Online]. Available: http://www.czso.cz/eng/redakce.nsf/i/living_conditions_household_income_and_expenditure_stat.
[3] Kawachi, I., Kennedy, P. B., Lochner, K. and D. Prothrow-Stith: Social Cupital, Income Inequality, and Mortality. *American Journal of Public Health* **87** (9), 1997, 1491 – 1498.
[4] Kuznets, I.: Economic Growth and Income Inequality. *American Economic Review* **65** (1), 1955, 1 – 28.
[5] Litchfield, J. A: Inequality: Methods and Tools. *Text for World Bank's Web Site on Inequality, Poverty, and Socio-economic Performance*, 1999. [Online]. Avaiable: http://www.worldbank.org/poverty/inequal/index.htm.
[6] Ministry of Finance: Údaje o vývoji daně z příjmů fyzických osob. 2014. [Online]. Available: http://www.mfcr.cz/cs/verejny-sektor/regulace/dane/danova/legislativa.
[7] Ministry of Labour and Social Affairs: Přehled o vývoji částek minimální mzdy. 2014. [Online]. Available: http://www.mpsv.cz/cs/871.
[8] Schutz, R.: On the Measurement of Income Inequality. *The American Economic Review* **41** (1), 1951, 107 – 122.
[9] Turečková, K.: Příjmové nerovnosti a jejich měření. *Acta Academica Karviniensia* **1** (2007), 191 -198.

Nonlinear DSGE model of a small open economy with time-varying parameters: Czech economy in a period of recession

Stanislav Tvrz[1], Osvald Vašíček[2]

Abstract. In this paper, we study the changes in the structure and behaviour of the Czech economy in a period of the Great Recession and subsequent return to the long-run growth equilibrium. This small open economy is represented by nonlinear dynamic stochastic model of a general equilibrium with financial accelerator. The development of time-varying structural parameters is identified using the second order approximation of a nonlinear DSGE model. The model is estimated with the use of nonlinear particle filter. Analogous model was estimated for the economy of the euro area. It is our goal to identify the most important changes in behaviour and underlying structure of the Czech economy. In order to distinguish the country specific changes from Europe-wide trends we also compare the time-varying estimates of the Czech economy and the euro area. We focus on the subset of parameters that are generally considered deep, i.e. time-invariant. Changes of these parameters during the period of global recession would imply that the structure of the economy together with its behaviour changed as a consequence of the recession.

Keywords: DSGE model, Great Recession, deep parameters, time-varying parameters, nonlinear particle filter, unscented particle filter.

JEL classification: E32, E44, E58
AMS classification: 91B64

1 Introduction

Between 2008Q4 and 2009Q2, in a period of global economic crisis also known as the Great Recession, the Czech economy experienced the biggest contraction in recent history when the real GDP declined by approximately 5 %. After two years of fragile expansion between 2009Q3 and 2011Q3, the real GDP almost returned to its pre-crisis maximum, however, another contraction followed. Between 2011Q4 and 2013Q1 the Czech economy experienced the longest recession in its history and the real GDP declined again by approximately 2.75 %.

Due to the unprecedented development of the Czech economy in last years, we find it reasonable to investigate the question of possible structural changes of the Czech economy. The aim of this paper is to identify the most important changes of the structural parameters during and after the Great Recession and to interpret these changes in terms of behaviour of representative economic agents. We also seek to find out to what extent did the structural changes of the Czech economy follow the common trends in the euro area and what were its unparalleled specifics.

In this paper, we estimate a DSGE model of the Czech small open economy with time-varying parameters and an analogous model of the economy of the euro area. We perform a filtration of a nonlinear model with the use of unscented particle filter (UPF) to identify the unobserved trajectories of the time-varying structural parameters.

We follow the direction of research of structural stability of the Czech economy of Vašíček *et al.* [6], who found evidence of parameter drifting in the Czech data. Also, we continue and elaborate on the research agenda introduced in Tvrz and Vašíček [4].

[1]Faculty of Economics and Administration, Lipová 41a, Brno, Czech republic, tvrz@mail.muni.cz
[2]Faculty of Economics and Administration, Lipová 41a, Brno, Czech republic, osvald@econ.muni.cz

2 The model

Since we focus on the period of financial and subsequent economic crisis, a DSGE model with financial frictions is used for the analysis. In our paper we use the model framework developed by Shaari [2] that includes financial accelerator mechanism proposed by Bernanke *et al.* [1]. This appropriately complex medium-sized model of a small open economy incorporates important real as well as nominal rigidities and allows us to describe the Czech economy in a reasonable detail. Structure of the model is quite standard, therefore, we will describe only the most important features of the model.

2.1 Households

The households receive wages for supplied labour, government transfers, profits made by retailers and domestic and foreign bonds returns. Domestic bonds pay fixed nominal return in domestic currency while foreign non-contingent bonds give a risk adjusted nominal return denominated in foreign currency. The debt-elastic risk premium contains exogenous AR(1) component of risk-premium or uncovered interest parity shock. The households then spend their earnings on consumption and domestic and foreign bonds acquisition.

2.2 Entrepreneurs

The entrepreneurs play two important roles in the model. They run wholesale goods producing firms and they produce and own the capital. Market of intermediate goods as well as capital goods market is assumed to be competitive. The wholesale goods production is affected by domestic productivity AR(1) shock and the capital goods production is subject to capital adjustment costs. Entrepreneurs finance the production and ownership of capital K_t by their net-worth N_t and borrowed funds. Cost of borrowed funds is influenced by borrower's leverage ratio via external finance premium,

$$EFP_t = \left(\frac{N_t}{Q_{t-1}K_t}\right)^{-\chi} \tag{1}$$

where Q_t is a real price of capital or Tobin's Q and χ is financial accelerator parameter. To maximize profit, the entrepreneurs choose the optimal level of capital and borrowed funds. Each turn a proportion $(1-A_t^{NW})\varsigma$ of entrepreneurs leaves the market and their equity $(1-A_t^{NW})\varsigma V_t$ is transferred to households in a form of transfers. A_t^{NW} is a shock in entrepreneurial net worth. It influences the development of net worth by changing the bankruptcy rate of entrepreneurs and its positive innovations increase the survival rate of entrepreneurs. Its logarithmic deviation from steady state is assumed to evolve according to AR(1) process. ς is the steady-state bankruptcy rate.

2.3 Retailers

Next, there are two types of retailers in the model. Domestic goods retailers and foreign goods retailers. Both types of retailers are assumed to operate in the conditions of monopolistic competition. Domestic good retailers buy domestic intermediate goods at wholesale price and sell the final domestic goods to the consumers. Foreign good retailers buy goods from foreign producers at the wholesale price and resell the foreign goods to the domestic consumers. The difference between foreign wholesale price expressed in domestic currency and final foreign goods price, i.e. deviation from law of one price is determined by exogenous AR(1) shock. By Calvo-type price setting and inflation indexation of the retailers the nominal rigidities are introduced into the model.

2.4 Central bank

The central bank determines the nominal interest rate in accordance with forward/backward-looking Henderson McKibbin Taylor interest rate rule. Deviations of interest rate from the interest rate rule are explained as monetary policy i.i.d. shocks.

2.5 Foreign sector

The foreign economy variables - real output, CPI inflation and nominal interest rate, are modelled using a structural VAR(1) model.

2.6 Time-varying parameters

All the model parameters are considered time-varying with the exception of shock autoregression parameters and standard deviations. The time-varying parameters are defined as unobserved endogenous variables with following law of motion

$$\theta_t = (1 - \alpha_t^\theta) \cdot \theta_{t-1} + \alpha_t^\theta \cdot \overline{\theta} + \nu_t^\theta \tag{2}$$

where θ_t is a general time-varying parameter, $\overline{\theta}$ is initial value of this parameter, α_t^θ is a time-varying adhesion parameter common for all the remaining time-varying parameters and $\nu_t^\theta \sim N(0, \sigma_\nu^\theta)$ is exogenous innovation in the value of parameter θ_t. Setting of the adhesion parameter α_t^θ influences the tendency of the time-varying parameter θ_t to return to its initial value $\overline{\theta}$. With $\alpha^\theta = 0$, the time-varying parameter would be defined as random walk, while with $\alpha^\theta = 1$, the parameter would be white noise centred around the initial value $\overline{\theta}$. For the purposes of this exercise, we set the initial value of the adhesion parameter to a value of $\alpha_0^\theta = 0.25$.

3 Estimation technique

Unscented particle filter (UPF) is used to identify the unobserved states of the DSGE model, including the time-varying parameters, in this paper. In this section, we briefly describe the main principles of this nonlinear particle filter.

3.1 Unscented particle filter

Unlike basic Kalman filter that is optimal only for linear systems with Gaussian noise, the unscented particle filter is a more sophisticated tool that can be used even for nonlinear state-space systems with non-Gaussian noise. In this section, we provide only the basic principles of the algorithm. A detailed description can be found for example in Van Der Merwe *et al.* [5]. In a condensed form, the UPF algorithm can be described as follows:

1. Initialization: $t = 0$, set the prior mean $\overline{x}_0$ and covariance matrix P_0 for the state vector x_t.
2. Generating particles: Draw a total of N particles $x_t^{(i)}$, $i = 1, \ldots, N$ from distribution $p(x_t)$ with mean $\overline{x}_t$ and covariance matrix P_t.
3. Time Update: $t = t + 1$, for each particle $(i = 1, \ldots, N)$ propagate the particle into future with the use of nonlinear transition and measurement equation and calculate means $\overline{x}_{t|t-1}$, $\overline{y}_{t|t-1}$ and covariance matrices $P_{t|t-1}$, $P_{y,y}$ and $P_{x,y}$.
4. Unscented Kalman filter: Calculate Kalman gain $K_t = P_{x,y} \left(P_{y,y}\right)^{-1}$, $\overline{x}_t = \overline{x}_{t|t-1} + K_t(y_t - \overline{y}_{t|t-1})$ and $P_t = P_{t|t-1} - K_t P_{y,y} \left(K_t\right)^T$
5. Continue by step 2 until $t = t_{max}$.

In our application we performed 20 runs of the UPF with 30.000 particles each for the second order approximation of the nonlinear DSGE model.

3.2 Initial values

Before the application of the UPF algorithm we estimated the model with constant parameters to obtain the estimates of autoregression parameters and standard deviations of structural shocks that are considered constant even in the UPF. Furthermore, the posterior means of the structural parameters were used as initial values of the time-varying parameters ($\overline{\theta}$). The standard deviations of time-varying parameters innovations (σ_ν^θ) were set equal to the 10 % of the posterior means of the structural parameters. Constant model parameters were estimated using Random Walk Metropolis-Hastings algorithm as implemented in Dynare toolbox for Matlab. Two parallel chains of 1.000.000 draws each were generated during the estimation. First 50 % of draws were discarded as a burn-in sample. The scale parameter was set to achieve acceptance rate around 30 %. Estimation of the model with constant parameters is described in detail in Tvrz *et al.* [3].

4 Data

Quarterly time series of eight observables were used for the purposes of estimation. These time series cover the period between 1999Q2 and 2013Q4 and contain 59 observations each.

Time series of real gross domestic product (GDP), harmonised consumer price index (CPI), 3-month policy interest rate and real investment are used for the domestic economy. The foreign economy is represented by the 17 euro area countries in the case of the Czech economy while the rest of the world is represented by the economy of the United States in the case of the euro area. Seasonally adjusted time series of real GDP, CPI and 3-month policy interest rate are used. Time series of CZK/EUR and EUR/USD real exchange rates are also used. These seasonally adjusted time series were obtained from the Eurostat, Czech National Bank, European Central Bank and Federal Reserve Bank of St. Louis databases.

The original time series were transformed prior to estimation so as to express the logarithmic deviations from their respective steady states. Logarithmic deviations of most observables from their trends were calculated with the use of Hodrick-Prescott filter ($\lambda = 1600$). Time series of the domestic and foreign CPI inflation were demeaned.

5 Empirical results

In this section, we present and discuss the most interesting empirical results. Table 1 contains the priors and posteriors of the parameters of the model with constant parameters estimated using the dataset of the Czech economy and of the economy of the euro area. Prior densities are the same for both economies in order to identify the structural differences in the data. Obtained results were used to calibrate the particle filter. The differences in the posterior estimates of the Czech economy and the economy of the euro area are intuitive and seem to capture the specifics of the two economies.

Figure 1 contains the trajectories of selected time-varying parameters.[1] Some parameters that were estimated as time-varying showed only negligible deviations from their initial values, and therefore, their trajectories are not presented. These parameters are namely the inverse elasticity of labour supply Ψ, the debt-elastic risk premium elasticity ψ_B, the price indexation κ and also the bankruptcy rate ς and the output gap weight in the Taylor rule Θ_y.

Most of the financial sector parameters showed significant deviations from the initial values in the period of 2008–2010. In the period of economic boom of 2008, the entrepreneurial net worth increased, which led to lower interest rate spreads. According to the trajectory of external finance premium elasticity χ, the interest rate spreads also became less sensitive to the variations in the leverage ratio. Improving situation of the entrepreneurs is also reflected in the trajectory of the steady-state leverage ratio Γ that decreased at that time, which means that the firms were becoming less dependent on the external financing. However, as the capital stock deviated above its steady state it became increasingly difficult to find suitable investment opportunities and capital adjustment costs Ψ_I increased. Steady-state bankruptcy

[1] The deviations from the initial values are expressed in per cent. Solid black line represents the results of the Czech economy while the dashed grey line corresponds to the results of the euro area. Vertical line denotes the period of 2009Q1 when the economic crisis in the Czech economy culminated.

rate showed only slight increase in the second half of 2008 in the Czech economy and later on in the euro area as well. The onset of the crisis in 2009 meant a correction and the values of the financial sector parameters slowly returned to the vicinity of their initial values.

			Prior		CZ Posterior		EA Posterior	
Parameter		**Distribution**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
Υ	Consumption habit	Beta	0.60	0.05	0.6005	0.0507	0.6716	0.0605
Ψ	Inv. elast. of lab. supply	Gamma	2.00	0.50	1.3439	0.3545	0.9513	0.2609
ψ^B	Debt-elastic risk premium	Gamma	0.05	0.02	0.0256	0.0077	0.0206	0.0074
η	Domestic/foreign elast. subst.	Gamma	0.65	0.10	0.5465	0.0766	0.4337	0.0211
κ	Price indexation	Beta	0.50	0.10	0.4951	0.0928	0.4349	0.0873
γ	foreign goods preference bias	Beta	0.40	0.15	0.4652	0.0692	0.2779	0.0358
θ_H	Domestic goods Calvo	Beta	0.70	0.10	0.8130	0.0281	0.7985	0.0261
θ_F	Foreign goods Calvo	Beta	0.70	0.10	0.8344	0.0246	0.8124	0.0271
ψ^I	Capital adjustment costs	Gamma	8.00	3.00	12.447	2.9236	16.177	3.3514
Financial frictions								
Γ	Steady-state leverage ratio	Gamma*	2.00	0.50	1.5076	0.2442	1.3060	0.2148
ς	Steady-state bankruptcy rate	Beta	0.025	0.015	0.0560	0.0238	0.0287	0.0137
χ	Financial accelerator	Gamma	0.05	0.015	0.0456	0.0135	0.0436	0.0131
Taylor rule								
ρ	Interest rate smoothing	Beta	0.70	0.10	0.8529	0.0218	0.7350	0.0358
β_π	Inflation weight	Gamma*	1.50	0.20	1.8311	0.2279	1.8491	0.2255
Θ_y	Output gap weight	Gamma	0.50	0.20	0.1768	0.0508	0.2277	0.0631

*Shifted gamma distribution is used for these parameters, because they are assumed to take values from $\langle 1, \infty)$.

Table 1 Estimated structural parameters

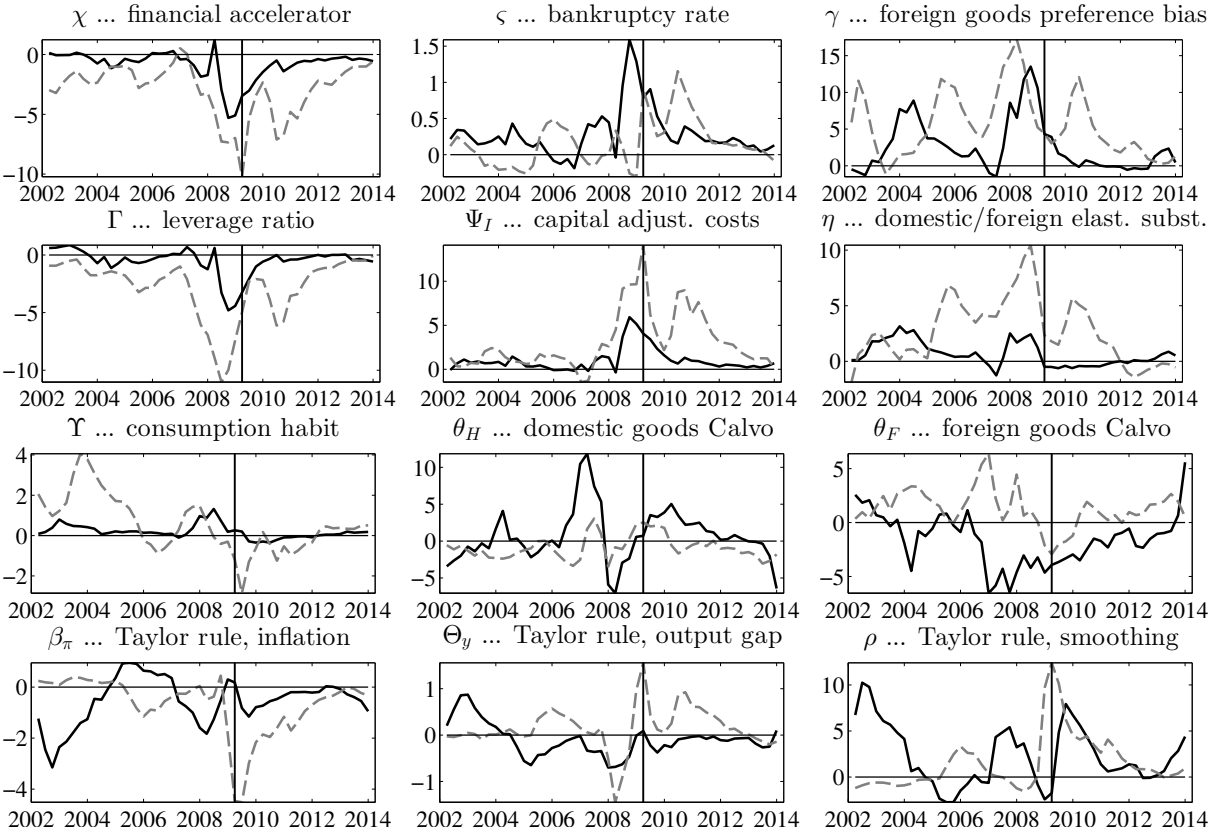


Figure 1 Selected time-varying parameters (deviations from initial values in per cent)

Apart from the financial parameters, there are other time-varying parameters that should not be omitted. Foreign goods preference bias parameter γ can be related to the openness of the economy and it shows a substantial increase in the period of 2007–2008 and subsequent return to the initial value after the 2009 crisis. Elasticity of substitution between domestic and foreign goods η deviated from the initial value in this period as well. Otherwise stable trajectory of the consumption habit Υ shows slight increase in the period of European Union enlargement of 2004 in case of the euro area economy. Development of the domestic goods Calvo parameter θ_H corresponds to the development of the domestic CPI inflation. According to the trajectory of the weight of inflation in the Taylor rule β_π, the lowering of the ECB's policy interest rates in 2009 was somewhat slower than would correspond to the initial value of this parameter. In case of the Czech economy, the inflation weight did not change considerably during the crisis. The smoothing parameter ρ is identified as the most time-varying among the parameters of the Taylor rule. A distinct increase of interest rate smoothing is apparent in the period of 2009–2010.

6 Conclusion

In this paper, we estimated two DSGE models of a small open economy with financial accelerator for the Czech economy and for the euro area. We applied a two-step approach to the estimation of the time-varying parameters. First, we estimated the models with time-invariant parameters using the Random Walk Metropolis-Hastings algorithm and then employed the obtained results for the initial setting of the unscented particle filter that was used for the estimation of the models with time-varying parameters. Some parameters showed only negligible deviations from their initial values, however, the parameters of the financial sector, openness parameter, elasticity of substitution between foreign and domestic goods, Calvo parameters and interest rate smoothing parameter changed markedly during the recent economic crisis of 2008–2009. Overall, the estimated trajectories of the time-varying parameters show many similarities between the development in the Czech economy and in the euro area with some differences in the magnitude of the deviations and timing. The differences can be attributed to earlier onset and more dramatic course of the financial crisis in the euro area than in relatively sheltered Czech economy.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0808/2013.

References

[1] Bernanke, B. S., Gerlter, M., and Gilchrist, S.: *The Financial Accelerator in a Quantitative Business Cycle Framework.* In: Taylor, J.B., and Woodford, M (eds.) Handbook of Macroeconomics, vol. 1, Elsevier, Amsterdam, 1999, 1341–1393.

[2] Shaari, M. H.: *Analysing Bank Negara Malaysia's Behaviour in Formulating Monetary Policy: An Empirical Approach.* Doctoral thesis, College of Business and Economics, The Australian National University, 2008.

[3] Tvrz, S., Tonner, J., and Vašíček, O.: *Financial Accelerator Mechanism in a Small Open Economy: DSGE Model of the Czech Economy.* In: Ramík, J., and Stavárek, D. (eds.). Proceedings of 30th International Conference Mathematical Methods in Economics, Karviná: Silesian University, School of Business, 2012, 944–949.

[4] Tvrz, S., and Vašíček, O.: *Nonlinear DSGE model of the Czech economy with time-varying parameters.* In: Vojackova, H. (ed.) Proceedings of the 31st International Conference Mathematical Methods in Economics. Jihlava: College of Polytechnics Jihlava, 2013, 968–973.

[5] Van Der Merwe, R., Doucet, A., De Freitas, N., and Wan, E.: *The Unscented Particle Filter.* CUED Technical Report 380, Cambridge University Engineering Department, 2000.

[6] Vašíček, O., Tonner, J., and Polanský, J.: *Parameter Drifting in a DSGE Model Estimated on Czech Data.* Finance a úvěr – Czech Journal of Economics and Finance, vol. 61, no. 5, UK FSV, Praha, 2011, 510–524.

Identification of player types in ultimatum bargaining games

Štěpán Veselý[1,2]

Abstract. Using cluster analysis I identified three types of players in two versions of the ultimatum bargaining game (with symmetric and with asymmetric information availability). I call the player types Fair players, Moderate opportunists and Hard-nosed opportunists. The player types constitute 23.1%, 43.3% and 33.6%, respectively, of the sample of 134 experimental participants. The types differ only moderately with respect to their choices in the standard ultimatum game. In particular, Hard-nosed opportunists make less generous offers than the two other types. However, there are very pronounced differences in behavior in the asymmetric information game. Fair players make fairer offers, misrepresent the surplus less frequently and to a lesser degree than the two other types. Hard-nosed opportunists are in turn less fair and more deceptive than Moderate opportunists in the asymmetric information game. Moreover, it is possible to infer that Opportunistic types are more strongly driven by fear of rejection (strategic concerns) and have weaker fairness preferencess than Fair types.

Keywords: Player types, cluster analysis, ultimatum game, deception, fairness.

JEL Classification: C72, C78

AMS Classification: 62H30

1 Introduction and theoretical background

Identification of player types and assessment of their prevalence constitutes an important problem in economics. Formulating realistic economic models (e.g. [3], [10]) can prove difficult, if not impossible without an accurate knowledge of the prevalence of specific player types.

Player types can be classified with respect to various criteria, such as their distributive preferences (e.g. [1]), propensity to punish social norms violation (e.g. [9]), fairness views (e.g. [4]) or strategic sophistication (e.g. [17]).

In this paper I classify participants in a bargaining experiment into player types according to two broad criteria: fairness of their choices and their tendency to deceive the other player when given an opportunity (these criteria are operationalized using four variables described in the Method section).

My analysis is based on data from a previous experiment that studied deception and fairness in ultimatum bargaining games (UGs) with symmetric and with asymmetric information [20].

UG with symmetric information is the standard UG introduced by [13]. It is a game for two players, usually with only one round. The players have a sum of money to be divided between them. One of the players is the proposer, the other is the responder. The roles are assigned randomly. The proposer makes an offer how to divide the money and the responder either accepts or rejects the offer. If the offer is accepted by the responder, the money is split according to the proposer's offer. However, if the responder rejects the offer, neither player gets anything. In both cases, the game is over. In the asymmetric information UG responders do not know the size of the surplus to be divided and proposers can give them false information about it (see e.g. [2], [15], [16] for studies using asymmetric information UG).

This paper contributes to the literature in that it classifies player types jointly according to two broad dimensions – their fairness and their tendency towards deception3. While several papers (briefly reviewed below) explicitly or implicitly sorted players into types according to each dimension separately, no previous paper provides a categorization that takes both dimensions jointly into account.

[1] Faculty of Arts, Masaryk University, Department of Psychology, Arne Nováka 1, Brno, Czech Republic e-mail: stepan.vesely@seznam.cz.
[2] Faculty of Business and Management, Brno University of Technology, Department of Economics, Kolejní 4, Brno, Czech Republic.
[3] More precisely, the classification is made according to four variables presented in the Method section (two measures of fairness and two measures of deception).

Several papers enable differentiation of players who deceive their counterparts and those who are honest. Note, however, that the prevalence of player types seems to depend, among other things, on the payoff structure, as well as on the absolute magnitude of rewards (see e.g. [12]).

For instance, in [15] proposers in a modified UG actively deceived in roughly 52% of cases. [16] found that 67%-89% of subjects (depending on the specific condition) sent untrue messages. In [2] the proportion of stated false surplus sizes was approximately 89%. In [20] 96% of subjects deceived at least once and 43% deceived whenever given an opportunity. Other examples include [5], [8], [12] who employed the sender-receiver game.

Cluster analysis is a convenient method for classification of objects according to one or several variables. Generally speaking, objects that are similar to each other are grouped into a common cluster, while they should be distinct from members of other clusters. There are many different clustering methods, see e.g. [14]. In this paper I use the two-step cluster algorithm by [6], which can be run in SPSS.

Cluster analysis is a method commonly used in economics (see e.g. [11], [19]). As far as I know, however, it has not been applied to classify players of economic games yet.

2 Method

2.1 Experimental design

The experiment is described in detail in [20]. Due to space limitations only the basic features are outlined here.

A total of 134 subjects played six one-round UGs. Players were rematched in each game. The size of the surplus divided in each game ranged between 20 and 200 units of experimental crowns (EC). The possible range of the surplus size was known to subjects in both experimental conditions, the probability distribution over this range that determined the amount of money to be divided was not known to participants.

The experiment had two stages. In stage 1, which is the one relevant here, all participants had the role of proposer. No feedback with respect to the acceptance of the proposals was given to them until after the experiment was completed. The offers from stage 1 were randomly distributed among participants in stage 2 to either accept, or reject them.

In stage 1 participants in the role of proposers made their decisions in six games: in three games in the symmetric information condition and in three games in the asymmetric information condition. In the symmetric condition the actual surplus size was known to both proposer and responder. In the asymmetric condition, the actual surplus size was known to the proposer, not to the responder. The proposer knew that the responder would not be given information about the actual surplus size. Instead, the surplus size stated by the proposer in stage 1 was announced by the experimenter as the actual surplus size to the responder in stage 2. Therefore, in asymmetric condition the proposer decided upon his/her division of the surplus and upon the size of the surplus announced to the responder in stage 2 as the actual surplus (this was the opportunity for active deception on the proposer's part).

2.2 Measures

Fairness of players' choices and their deceptive behavior are operationalized using the following four variables: fairness of their choices in symmetric information UG (variable FAIR_SI), fairness of their choices in asymmetric information UG (variable FAIR_AI), relative frequency of deception in asymmetric information UG (variable DEC_FR), and degree of deception in asymmetric information UG (variable DEC_DEG).

FAIR_SI is calculated as the percentage of surplus offered to the responder averaged over the three decisions each proposer made in symmetric information UG.

FAIR_AI is calculated as the percentage of (real) surplus offered to the responder averaged over the three decisons each proposer made in asymmetric information UG.

DECEP_FR is calculated as the number of games where deception was present divided by the total number of asymmetric information games (i.e., three). Cases of actual surplus equal to 20 EC are ignored, because deception has no sense in these cases (see Veselý, 2014).

DECEP_DEG equals one minus stated surplus size divided by actual surplus size. This measure is again averaged over the three decisons each proposer made in asymmetric information UG. Cases of actual surplus equal to 20 EC are ignored.

3 Results and discussion

Using two-step cluster algorithm with Schwarz's Bayesian Criterion I obtained three clusters of players. The silhouette measure of cohesion and separation of this three-cluster solution is fairly good (above 0.5). Table 1 reports the cluster profiles – means (and standard deviations) for the three clusters. Each cluster profile povides a characterization of a player type. The first type is Fair type, the second type is Moderate opportunist, and the third type is Hard-nosed opportunist.

Cluster / Player type	FAIR_SI	FAIR_AI	DECEP_FR	DECEP_DEG
Fair players	.50 (.08)	.45 (.04)	.39 (.24)	.11 (.08)
Moderate opportunists	.46 (.04)	.33 (.05)	.79 (.16)	.35 (.10)
Hard-nosed opportunists	.42 (.10)	.19 (.06)	.98 (.08)	.63 (.12)

Note: The table gives mean and (SD) for each variable by type / cluster.

Table 1 Profiles of three player types / clusters of players

As can be seen from Table 1, Moderate opportunists fall somewhere between Fair players and Hard-nosed opportunists in all measured variables. There are 31 Fair players (23.1% of the sample), 58 Moderate opportunists (43.3% of total) and 45 Hard-nosed opportunists (33.6% of total).

Fair players seem to possess genuine preferences for fairness, rather than to be driven by fear of rejection when making their offers (see Suleiman, 1996; Veselý, 2014). They share roughly half of the surplus with the responder irrespective of the information condition (see the second and third column in Table 1). However, even this type is driven by fear of rejection, because FAIR_AI (M = .45) is significantly smaller than FAIR_SI (M = .50) according to a paired t-test: $t(30) = 3.44$, $p < .01$, effect size $r = .53$. The effect size is quite large, in fact.

For Moderate and Hard-nosed opportunists, the effect size of information condition is larger, which means they are more strongly driven by fear of rejection and have weaker fairness preferencess than Fair players. In particular, for Moderate opportunists, the difference between fairness of offers in symmetric (M = .46) and asymmetric information UG (M = .33) is significant according to a paired t-test: $t(57) = 14.35$, $p < .001$, effect size $r = .88$. The effect size is very large. The difference between fairness of offers in symmetric (M = .42) and asymmetric information UG (M = .19) is about as much pronounced for Hard-nosed opportunists: $t(44) = 11.88$, $p < .001$, effect size $r = .87$ (again, a very large effect).

Fair players misrepresent the surplus in somewhat over one case in three on average and on average, they misrepresent the surplus by 11% (DECEP_DEG = .11). In contrast, Hard-nosed opportunists lie basically every time (DECEP_FR = .98). Moderate opportunists lie in roughly 80% of cases, i.e. also quite a lot. However, Moderate opportunists misrepresent the surplus much less than Hard-nosed opportunists. DECEP_DEG = .35 on average in case of Moderate opportunists, whereas in case of Hard-nosed opportunists it is .63 on average, the difference is statistically significant accordng to independent samples t-test: $t(101) = 13.60$, $p < .001$, effect size $r = .80$, large effect.

Quite fascinating is the fact that players of different types behave relatively similarly in the standard UG (compare FAIR_SI by player type in Table 1). To see whether the differences in FAIR_SI by player type are significant, I ran a one-way ANOVA with player type as an independent variable and FAIR_SI as an outcome variable. There was in fact an overall significant difference among player types with respect to their behavior in the standard UG, $F(2, 131) = 11.90$, $p < .001$, the effect size $\omega = .37$ was of medium size. Using post hoc tests to compare player types pair-wise shows there are significant differences between Hard-nosed opportunists and both other types ($p < .01$ in both cases), but only a marginally significant difference between Fair players and Moderate opportunists ($p < .1$ when using Tukey HSD, Bonferroni, Gabriel).

As a comparison, consider the differences among player types with respect to their offers in the asymmetric information UG (FAIR_AI) and with respect to deception (DECEP_DEG). These differences were also significant ($p < .001$ in both cases), but the effect sizes were much larger, $\omega = .89$ both for FAIR_AI and DECEP_DEG. This shows that behavior across player types was only moderately different in the standard UG, but extremely different in the asymmetric information UG.

Motivations underlying behaviors in the standard UG are presumably rather different for different types. Namely, Fair players are motivated mostly by their fairness preferences, whereas opportunistic types are

motivated to a larger degree by fear of rejection This can be inferred from the comparisons between FAIR_SI and FAIR_AI that I performed above, for general discussion see e.g. [18] and [20].

Using the research strategy of comparing players' behavior in symmetric and asymmetric UGs enables us to investigate whether players are motivated by fairness prefereces, or by fear of rejection (or strategic concerns, to put it differently). Unlike previous studies that focused on this issue (e.g. [2], [7], [15], [16]), the present paper also demonstrates differences in terms of behavior and likely motivations across three different types of players in ultimatum bargaining games.

4 Conclusion

Although empirical identification of player types in different strategic games is an impotant research objective in economics, the types are usually classified only along a single dimension (e.g. their level of strategic reasoning). In this paper I employed cluster analysis to classify agents interacting in two versions of the ultimatum game according to their fairness and deceptiveness. Three types emerged from the analysis, Fair players, Moderate opportunists and Hard-nosed opportunists. Fair players were fairer (more generous) and less deceptive than the Opportunistic types. Interestingly, differences among players were only moderate in the standard symmetric information ultimatum game, but they were very large in the asymmetric information game.

References

[1] Andreoni, J., and Miller, J.: Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* **70** (2002), 737–753.

[2] Besancenot, D., Dubart, D., and Vranceanu, R.: The value of lies in an ultimatum game with imperfect information. *Journal of Economic Behavior & Organization* **93** (2013), 239–247.

[3] Bolton, G. E., and Ockenfels, A.: ERC: A theory of equity, reciprocity, and competition. *American Economic Review* **90** (2000), 166–193.

[4] Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B.: The pluralism of fairness ideals: An experimental approach. *American Economic Review* **97** (2007), 818–827.

[5] Childs, J.: Gender differences in lying. *Economics Letters* **114** (2012), 147–149.

[6] Chiu, T., Fang, D., Chen, J., Wang, Y., and Jeris, C.: A robust and scalable clustering algorithm for mixed type attributes in large database environment. In: *Proceedings of the 7th ACM SIGKDD international conference on knowledge discovery and data mining*. Association for Computing Machinery, 2001, San Francisco, CA, 263–268.

[7] Croson, R. T. A.: Information in ultimatum games: An experimental study. *Journal of Economic Behavior & Organization* **30** (1996), 197–212.

[8] Dreber, A., and Johanesson, M.: Gender difference in deception. *Economics Letters* **99** (2008), 197–199.

[9] Fehr, E., and Fischbacher, I.: Third-party punishment and social norms. *Evolution and Human Behavior* **25** (2004), 63–87.

[10] Fehr, E., and Schmidt, K. M.: A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* **114** (1999), 817–868.

[11] Ferenčíková, S., Mühlbacher, J., Kodydek, G., and Nettekoven, M.: Role allocation and strategic change: Comparative study of competencies of Austrian and Slovak managers. *Ekonomický časopis* **60** (2012), 717–731.

[12] Gneezy, U.: Deception: The role of consequences. *American Economic Review* **95** (2005), 384–394.

[13] Güth, W., Schmittberger, R., and Schwarze, B.: An experimental analysis of ultimatum games. *Journal of Economic Behavior & Organization* **3** (1982), 367–388.

[14] Kaufman, L., and Rousseeuw, P. J.: *Finding groups in data: An introduction to cluster analysis*. Wiley, Hoboken, NY, 2005.

[15] Koning, L., Steinel, W., Van Veest, I., and Van Dijk, E.: Power and deception in ultimatum bargaining. *Organizational Behavior and Human Decision Processes* **115** (2011), 35–42.

[16] Kriss, P. H., Nagel, R., and Weber, R. A.: Implicit vs. explicit deception in ultimatum games with incomplete information. *Journal of Economic Behavior & Organization* **93** (2013), 337–346.

[17] Stahl, D. O., and Wilson, P. W.: On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior* **10** (1995), 218–254.

[18] Suleiman, R.: Expectations and fairness in a modified ultimatum game. *Journal of Economic Psychology* **17** (1996), 531–554.

[19] Tola , V., Lillo , F., Gallegati , M., and Mantegna, R.: Cluster analysis for portfolio optimization. *Journal of Economic Dynamics & Control* **32** (2008), 235–258.

[20] Veselý, Š.: Ultimatum game with asymmetric information: A study of deception and fairness. *Studia Psychologica* **56** (2014), 53–66.

On Bayes approach to optimization

Petr Volf [1]

Abstract. In many real optimization problems the objective function is either hardly tractable or its evaluation is expensive. Hence, we have not full information on its form and can afford to evaluate it at just a few points. Then, certain assumptions on the objective function form (shape) must be done. This could be with advantage taken as a prior information in a Bayes scheme. The Bayes approach to optimization, extensively studied in last several decades, then offers the way of effective search for the extremal point. In the present paper we shall recall the ideas behind Bayes optimization procedures, describe the technique using the model of Gauss process and derive a regression-like method dealing with noisy information on objective function.

Keywords: optimization, Bayes method, MCMC, Gauss process, nonparametric regression.

JEL classification: C41, J64
AMS classification: 62N02, 62P25

1 Introduction

The paper deals with the problem of optimization in the case that we have not complete information on the objective function and that its evaluation is costly (computationally or even literally). Hence, we wish to reduce its evaluation just to a few points. Therefore, we have to find an effective way how to ,,reconstruct" objective function, at least in a neighborhood of its extremal point. To make it possible, it is necessary to make some assumptions on the objective function form. These assumptions can as well be taken as a prior information, which is further specified on the basis of other reasonably selected observations. Such a point of view leads quite logically to the use of Bayes approach.

In the present paper two cases are distinguished. In Part 2, it is assumed that the objective function can be evaluated (practically) without error. Then, one of the convenient methods takes the form of objective function as a trajectory of Gauss process, with parameters describing its covariance structure. Such an approach is in detail described in Brochu et al (2010). The authors provide also a rich list of relevant references dated from 70-ties till today and covering also some competing approaches.

However, when the objective function is evaluated with random noise, the interpolating model of Gauss process has problems not only with tracing unknown function shape but also with locating its mode. Therefore in this case we prefer non-parametric regression model estimating the objective function and extrapolating its estimate to the region where the mode is expected. This approach is described in Part 3. The objective function is constructed from regression (i.e. smoothing) splines (De Boor, 1978). Their parameters, namely the location of splines knots and also their numbers, are subjected to bayesian estimation. Practical Bayes search then uses the MCMC (Markov chain Monte Carlo) procedures. More on the MCMC method can be found in a number of papers and monographs, see for instance Gamerman (1997) and also Volf (2006).

In both cases, the search procedure should be adaptive in the sense of a trade-off between *exploration* (revealing the objective function shape) and knowledge *exploitation*, especially in the neighborhood of expected extremal point. In practise it means that after each step the posterior is re-analyzed and the next step selected on the basis of re-computed predictive probabilities.

[1]Institute of Information Theory and Automation of the ASCR, Pod vodárenskou věží 4, Praha 8, Czech Republic, volf@utia.cas.cz

2 Gauss process model

The use of Gauss process as a model for unknown objective function dates back to 70-ties. The advantage is that all finite-dimensional and also conditional distributions are normal. Hence, the process is fully specified by its expectation function $m(z)$ and covariance function $\mathrm{cov}(z_1, z_2)$. In the present case we consider an isotropic (homogeneous) process with constant $m(z) = m$ and covariance depending only on distance $|z_1 - z_2|$. Namely,

$$\mathrm{cov}(z_1, z_2) = d^2 \cdot \exp(-c\,|z_1 - z_2|^2) \tag{1}$$

is a convenient choice (Brochu et al, 2010, Močkus, 1994). Thus, the Gauss process is described by three parameters $m, d > 0, c > 0$.

Let us assume that we already know values of objective function $g_i = g(z_i)$ at several points $z_i, i = 1, ..., n$. Our task is now to select another point z convenient for new evaluation of objective function, having in mind that we wish to approach the $\mathrm{argmax}\, g(z)$. Let $g(z)$ be the value at a new point z, further let us denote $\boldsymbol{m} = (m, ..., m)'$, $\boldsymbol{g} = (g_1, ..., g_n)'$, $\boldsymbol{D} = \mathrm{cov}(\boldsymbol{g})$, and $\boldsymbol{d}_z = \mathrm{cov}(\boldsymbol{g}, g(z))$. Then the joint distribution of ,,old" and new values is

$$\begin{pmatrix} \boldsymbol{g} \\ g(z) \end{pmatrix} \sim N\Big[\begin{pmatrix} \boldsymbol{m} \\ m \end{pmatrix}, \begin{pmatrix} \boldsymbol{D}, \boldsymbol{d}_z \\ \boldsymbol{d}_z', d^2 \end{pmatrix}\Big].$$

Finally, the conditional distribution of the interest is

$$(g(z) - m|\boldsymbol{g}, z) \sim N(\mu(z), \sigma^2(z)), \tag{2}$$

with

$$\mu(z) = \boldsymbol{d}_z'\, \boldsymbol{D}^{-1}\, (\boldsymbol{g} - \boldsymbol{m}), \quad \sigma^2(z) = d^2 - \boldsymbol{d}_z'\, \boldsymbol{D}^{-1}\, \boldsymbol{d}_z.$$

In Bayes setting, the prior space is the space of trajectories of the Gauss process, with its prior distribution given by parameters m, c, d. It means that they are taken as hyper-parameters. The likelihood is then given by the joint Gauss distribution of the data, which actually also stands for a finite-dimensional part of the Gauss process posterior. We are, however, mainly interested in the conditional distribution (2), as it represents the Bayes predictive distribution of the next observation. It is seen that when the hyper-parameters are selected (fixed), the Bayes ‚scent' is presented just implicitly, offering one of several possible interpretations of the method. The choice of covariance function (1) can as well be taken as a selection of a kernel influencing the smoothness of the Gauss process trajectories. Hence, some rules for the choice of optimal smoothing kernel can be applied (again, cf. Brochu et al, 2010).

The main aim of the model construction is to provide a tool for the selection of the next point which, with high probability, is closer to $\mathrm{argmax}\, g(z)$ than points already screened. Let $z^+ = \mathrm{argmax}\, g(z_i), i = 1, ..., n$ be the extremal point from them. Then the ,,probability of improvement", i.e. the probability that at a point z the value is higher, is

$$PI_0(z) := \mathrm{P}(g(z) > g(z^+)) = 1 - \Phi\Big(\frac{g(z^+) - \mu(z)}{\sigma(z)}\Big),$$

where $\Phi(.)$ denotes the distribution function of standard Gauss distribution. Intuitively we should select z maximizing $PI_0(z)$. However, this choice has tendency to dwell in the vicinity of z^+ not supporting jumps to other areas. Therefore, an improved rule is based on the criterion

$$PI(z) := \mathrm{P}(g(z) > g(z^+) + v) = 1 - \Phi\Big(\frac{g(z^+) + v - \mu(z)}{\sigma(z)}\Big),$$

(cf. Kushner and Yin, 1997, Torn and Žilinskas, 1989). Here $v \geq 0$ is a tuning parameter, chosen rather subjectively. It is recommended to decrease it to zero (exponentially, for instance) with growing number of procedure iterations. An analogy can be found in tuning the cooling parameter in the simulated annealing method of randomized optimization. The following simple example illustrates the method on the case of one-dimensional objective function, however, generalization to more dimensions is quite straightforward.

Example 1. Let us consider an objective function

$$g(z) = 10 - z \cdot \sin(\frac{z^2}{10})$$

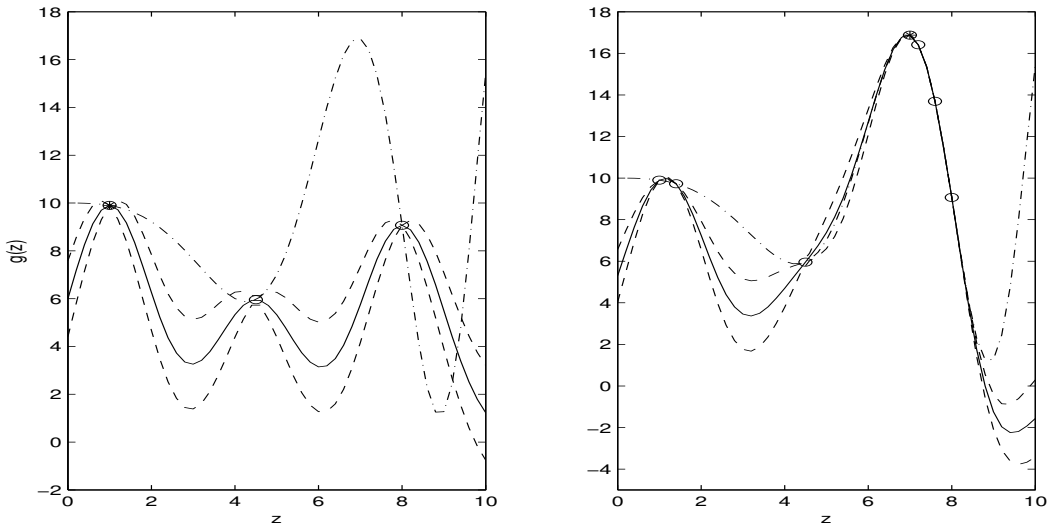


Figure 1 Use of Gauss process model: Initial phase of search with just 3 points (left), state of search after 4 iterations (right)

on interval $\boldsymbol{Z} = (0, 10)$ and let us search for its maximum. Further, let us assume that, initially, the function is evaluated at only 3 points. Figure 1, left plot, shows such a case, maximum of $g(z_i)$, $i = 1, 2, 3$, is at the left point. Solid curve shows the mean $\mu(z)$ of Gauss process constructed from these 3 points, while two dashed curves show $\mu(z) \pm 2\sigma(z)$. Dash-dots curve is then (unknown) objective function $g(z)$. The right plot shows the progress of search after 4 iterations of the search procedure described above. Function has been evaluated sequentially at 4 new points proposed by the PI. They are shown in the plot, the last of them is already quite close to the mode of $g(z)$. In this experiment the parameters of the procedure were fixed to $d = 1, c = 0.5$, m was estimated by the mean of $g(z_i)$. Tuning parameter in j-th iteration was set to $v(j) = 0.5^j$.

3 Nonparametric regression model

Let us now assume that evaluation of the objective function is not precise, that instead $g(z)$ we observe $y(z) = g(z) + \varepsilon(z)$. In the simplest case it is assumed that $\varepsilon(z)$ are independent copies of the same random variable ε possessing the Gauss distribution with zero mean and an unknown variance δ^2. There are essentially two different ways how to estimate unknown (we assume that smooth enough) function $g(z)$. The first consists in the local (e.g. kernel) smoothing. The other approach, utilized here, employs the approximation of $g(z)$ by a combination of functions from some functional basis. For instance, the polynomial splines are the popular choice. Then the model of function $g(z)$ has the form

$$g_M(z) = \boldsymbol{\alpha}' \boldsymbol{B}(z; \boldsymbol{\beta}) = \sum_{j=1}^{M} \alpha_j B_j(z; \boldsymbol{\beta}), \tag{3}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)'$ is a vector of linear parameters, B_j are basis functions and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)'$ is a vector of parameters of the basis functions (e. g. knots of splines). While the estimates of $\boldsymbol{\alpha}$ can be obtained directly from linear regression context, estimation of $\boldsymbol{\beta}$ is more difficult. As a solution to the nonlinear problem for coefficients $\boldsymbol{\beta}$ as well as to optimal choice of number of used units, M, it is possible to use the Bayes methodology in combination with the Markov chain Monte Carlo (MCMC) algorithms. In this framework, the parameter $\boldsymbol{\beta}$ is considered to be a multi-dimensional random vector, with a prior distribution satisfying certain constraint. Simultaneously, M is also regarded as a random variable, with a decreasing prior on $\{0, 1, 2, \ldots, M_{max}\}$. Such a choice lowers the chance to accept a model with high number of units, if the gain of that model is low.

Example 2. Let us consider the same objective function $g(z)$ as in Example 1. However, now it is assumed that it is evaluated with a random error, namely at selected z we observe

$$y(z) = g(z) + \varepsilon(z),$$

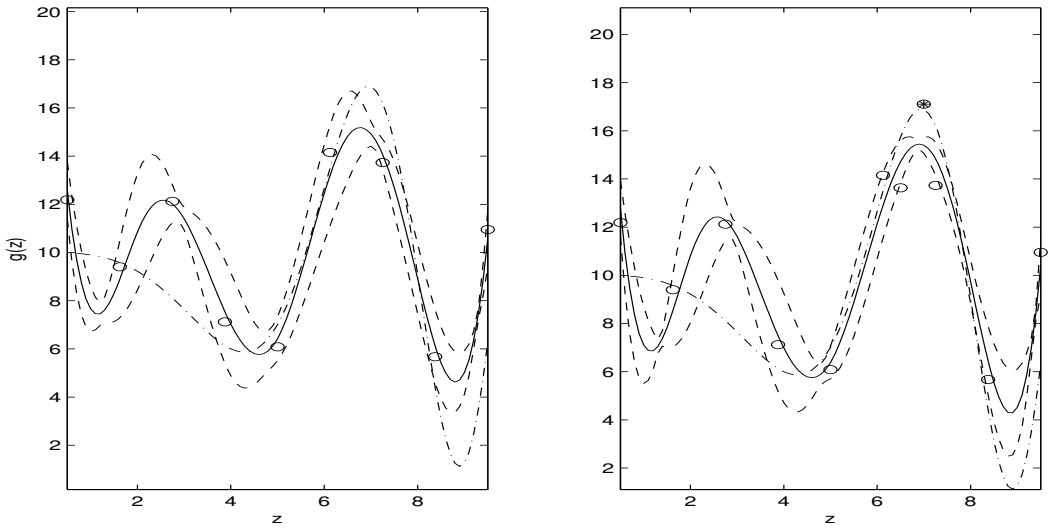


Figure 2 Use of splines model: Initial phase of search for maximum of objective function (left), state of search after 2 iterations (right)

where $\varepsilon(z)$ are the i.i.d. random variables $\varepsilon(z) \sim N(0, \delta^2 = 4)$.

The procedure started from observations at 7 points $z(i)$ located uniformly inside $(0, 10)$, where values $y(z_i) = g(z_i) + \varepsilon(z_i)$ were generated. Notice that here, in the case with considerable noise, the initial number of point is larger than in Example 1. It is necessary for evaluation of the splines.

Values $y(z_i)$ are shown in Figure 2, left plot, again, hidden objective function $g(z)$) is dashed. Estimate of $g(z)$ was then constructed from cubic B-splines. As regards the prior for their knots, we used uniform distribution on the set $\{0 < \beta_1 < \beta_2 < \ \ldots \ < \beta_M < 10\}$. M was bounded by 6 in order to ensure their identifiability. $S = 500$ loops of the Markov chain generation were performed. One loop updated sequentially all components of $\boldsymbol{\beta}$, with possible change of M. It means that it contains up to 6 iterations of model, depending on actual number M.

Only the final result after each loop was registered as a new member of the chain, $g^{(m)}(z)$. The average of this sequence of functions, after skipping first $s = 100$ of them,

$$\mu(z) = \frac{1}{S-s} \sum_{m=s+1}^{S} g^{(m)}(z), \tag{4}$$

serves then as the estimate of $g(z)$. In Figure 2 it is plotted by a full curve. The variability of this set of $S - s$ functions is not constant. The vertical cut at a given z represents Bayes prediction distribution for corresponding $g(z)$. Hence, variance of prediction $\sigma^2(z)$ is computed as a sample variance from values $g^{(m)}(z)$ (compare also discussion in Bishop, 1992, Ch. 10). Dashed curves in the plot again show $\mu(z) \pm 2\,\sigma(z)$. The right plot shows also two new sequentially chosen points tending to the mode of $g(z)$.

Computation of prediction variance

When the nonlinear part of the model (e.g. the knots of splines) is specified, the variance of prediction can also be quantified with the aid of standard linear regression analysis adapted to our case. We then deal with the following linear regression model

$$y_i = \boldsymbol{B}^T(x_i) \cdot \boldsymbol{\alpha} + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\boldsymbol{\alpha}$ are unknown parameters, $\varepsilon_i \sim \mathcal{N}(0, \delta^2)$ are the i.i.d. normal random variables, $\boldsymbol{B}(x_i) = (B_1(x_i), \ldots, B_M(x_i))^T$ are B-splines evaluated at data-points $\boldsymbol{x} = (x_1, \ldots, x_n)^T$. Denote $\boldsymbol{B}$ the $n \times M$ matrix with rows $\boldsymbol{B}^T(x_i)$, $\boldsymbol{A} = (\boldsymbol{B}^T \cdot \boldsymbol{B})^{-1}$, $\boldsymbol{y} = (y_1, \ldots, y_n)^T$. Then the least squares method yields the estimate

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{A} \cdot \boldsymbol{B}^T \cdot \boldsymbol{y}, \quad \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \sim \mathcal{N}(O, \delta^2 \cdot \boldsymbol{A}),$$

where O is the null vector. Further, at a new point z the prediction of $g(z)$ is $\hat{g}(z) = \boldsymbol{B}^T(z) \cdot \hat{\boldsymbol{\alpha}}$. Its expectation is 'true' $g(z)$, while its variance equals

$$\sigma^2(z) = \text{var}\left(\hat{g}(z)\right) = \boldsymbol{B}^T(z) \cdot \boldsymbol{A} \cdot \boldsymbol{B}(z) \cdot \delta^2.$$

As expected, it depends both on data ($\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{x})$) from which the model was estimated, and on position of prediction point z.

A case with multidimensional objective function can be, essentially, solved in the same manner. However, multivariate function has to deal with interactions of several predictors. This is as a rule modelled by a tensor product of one-dimensional units. The problem is caused by the fact that their number grows (exponentially) with dimension, so that there also grows the number of 'nonlinear' parameters. Nevertheless, there are some approaches to the regression functions modelling able to reduce the number of parameters, for instance the projection pursuit method.

4 Application to quantile optimization problem

In quantile optimization the main criterion of interest is certain quantile related to the decision maker risk acceptance. Let us consider a stochastic optimization problem with utility function $\varphi(x, z)$, where z are input (decision) variables from a set $\boldsymbol{Z}$ and values x are results of a random variable (or vector) X. Denote $F(y; z)$ the distribution function of random variable $Y = \varphi(X, z)$ with decision z, then, for selected $\alpha \in (0, 1)$ the objective is to maximize, over $z \in \boldsymbol{Z}$, α-quantile $Q(\alpha; z) = \min\{y : F(y; z) \geq \alpha\}$. However, in many cases the evaluation and consequent optimization of quantile criterion is not easy, often it uses iterative procedures or Monte Carlo generation. Then the method described in the present paper can be applied. We shall show such an application of method described in Section 2 on an example from the area of reliability and maintenance optimization.

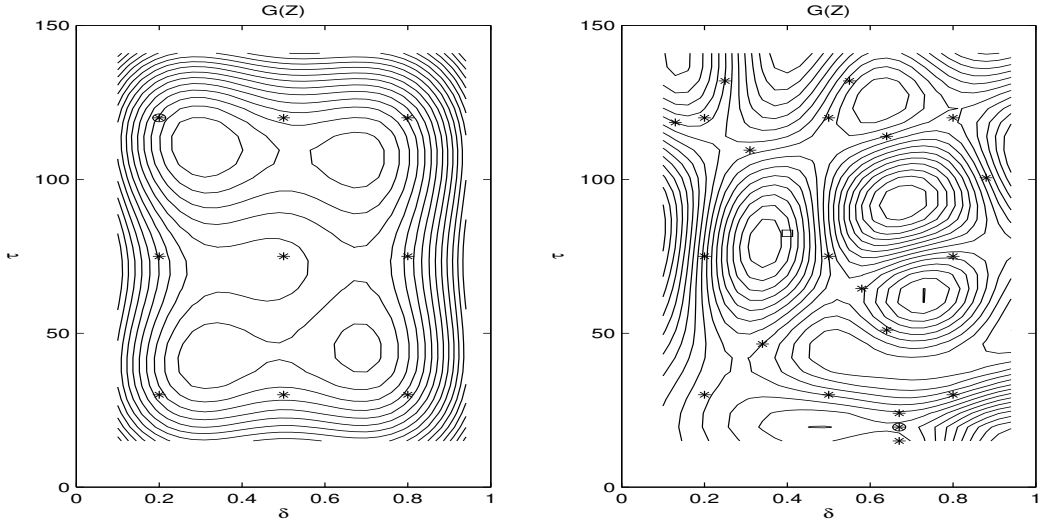


Figure 3 Example 3: Initial phase of search (left), state of search after 12 iterations (right)

Example 3. The Kijima model of non-complete repair (Kijima, 1989) assumes that the device is repaired in its age τ with a degree δ, which means that after repair the virtual age of the device is $(1 - \delta) \cdot \tau$. Thus, $\delta = 1$ means complete repair, renewal, while $\delta = 0$ is the minimal repair.

In the example it is assumed that the Kijima model concerns to preventive repairs, meanwhile after the failure the device has to be renewed completely. We are given the costs of renewal, C_1, and of preventive repair, $C_2(\delta, \tau)$. The objective is to maximize, over τ and δ, an α-quantile of random function $\varphi(X, \delta, \tau)$ equal to proportion of the time to renewal to the costs to renewal. Here X is the random time to failure of the device. This proportion equals

$$\varphi(X, \delta, \tau) = \frac{X}{C_1} \quad \text{with probability} \quad P(X \leq \tau),$$

$$\varphi(X,\delta,\tau)=\frac{\tau+\tau\cdot\delta\cdot(k-1)+X_k}{C_1+k\cdot C_2}\text{ with }P(X>\tau)\cdot P(X_1>\tau)^{k-1}\cdot P(X_k\le\tau),$$

where all $X_k=\{X|X>\tau(1-\delta)\}$ and k is the number of preventive repairs before the failure. It is seen that the direct evaluation of objective function is not easy, moreover, it is strongly non-concave. Therefore, the distribution of variable $Y(\delta,\tau)=\varphi(X,\delta,\tau)$, for different δ,τ, is obtained ‚empirically' by random generation, quantiles $Q(\alpha;\delta,\tau)$ then as sample quantiles.

For numerical illustration we selected $X\sim\text{Weibull}(a=100,b=2)$, with survival function $\overline{F}(x)=\exp\Big(-\big(\frac{x}{a}\big)^b\Big)$, E$X\sim 89$, std$(X)\sim 46$. Further, $\alpha=0.1$ the costs $C_1=40$, $C_2=2+(\delta\cdot\tau)^\gamma$, $\gamma=0.2$. Figure 3 shows the results. Objective function is $Q(\alpha;\delta,\tau)$, procedure started from its Monte Carlo generation in 9 points showed in the left plot. Maximum is denoted by a circle, its value was 0.876. The plot contains also contours of resulting Gauss process surface. The right plot shows the situation after 12 iterations. It is seen how the space was inspected, maximal value was stabilized around 1.124, the corresponding point ($\delta\sim 0.7,\tau\sim 20$) is again marked by a circle.

5 Conclusion

The contribution has studied the problem of optimization in the case when the objective function is not known sufficiently and its evaluation is costly. Two methods of the search for the objective function extremal point were described. Both are based on an appropriate model of the function. However, the aim is not to reconstruct it fully (like in the regression analysis), but, preferably, in the vicinity of its extreme, minimizing simultaneously number of function evaluations. The success of the methods is based also on the design of starting points. Like in other global optimization methods, if they are far from the optimum and the function is not unimodal, the procedure can end in a local extreme.

Acknowledgements

The research has been supported by the grant No. 13-14445S of the Czech Science Foundation.

References

[1] Bishop, C.: *Neural Networks for Pattern Recognition*. Cambridge Univ. Press, Cambridge, 1992.

[2] Brochu, E., Cora, V. M., and de Freitas, N.: A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv:1012.2599v1 [cs.LG]* **12** (2010).

[3] De Boor, C.: *A Practical Guide to Splines*. Springer Verlag, Berlin, 1978.

[4] Gamerman, D.: *Markov Chain Monte Carlo*. Chapman and Hall, New York, 1997.

[5] Kijima, M.: Some results for repairable systems with general repair. *Journal of Applied Probability* **26** (1989), 89–102.

[6] Kushner, H. J. and Yin, G. G.: *Stochastic Approximation Algorithms and Applications*. Springer Verlag, Berlin, 1997.

[7] Močkus, J.: Application of Bayesian approach to numerical methods of global and stochastic optimization. *J. Global Optimization* **4** (1994), 347–365.

[8] Torn, A. and Žilinskas, A.: *Global Optimization*. Springer Verlag, Berlin, 1989.

[9] Volf, P.: Markov Chain Monte Carlo methods in computational statistics and econometrics. In: *Proceedings of the 24th International Conference Mathematical Methods in Economics 2006* (L. Lukáš ed.), ZČU Plzeň, 2006, 525–530.

Modelling, Visualization and Optimization of Hospital Emergency Department

Jan Voracek[1], Hana Vojackova[1], Martina Kuncova[1], David Zazimal[2]

Abstract. Hospital emergency department is a complex job shop with advanced routing logic and nontrivially shared resources. Efficient management of such unit represents typical multicriterial decision making problem, balancing throughput of patients with operational costs and quality of care. Because of explicit structure and well defined processes, management of ED is good candidate for computational modelling and simulation with the discrete events paradigm. Proposed paper analyses, models and optimizes main clinical and administrative processes in ED in order to orchestrate their mutual interactions and suggest feasible performance improvements in terms of average length of stay, average waiting time, utilization of resources and total costs. Consequently, we propose several scenarios, demonstrating gradual parametric optimization of the initial overestimated and costly configuration. After its exemplary extension with the Theory of Constraints, Lean and Six Sigma principles we achieved comparable performance with considerably less resources. Finally we propose also elementary structural changes, leading to the best obtained solution.

Keywords: Discrete events simulations, Operations management, Emergency department, Lean, Six Sigma.

JEL Classification: C63
AMS Classification: 90B22, 90B50

1 Introduction

This paper proposes a generalized simulation model of hospital emergency department (ED). The main goal of our research was to demonstrate managerial usefulness of discrete events (DE) modelling and straightforward applicability of simulated results. Consequently, special emphasis was given to the discussion of results with respect to standardized management frameworks, as well as to the realistic visualization of clinical operations. ED is a front end medical unit, serving as a fast entry point to hospital for different types of urgent patients. Inpatients are primarily classified according to severity of their problems as follows:

- Immediate and emergent patients, who must be processed as fast as possible. In our case, they create 15% of all patients, where 13% of them enter ED by ambulance (A) and 2% by foot (F)),
- Urgent patients, who can wait up to tens of minutes (51%, where A = 4% and F = 47%),
- Semi-urgent and non-urgent, who are handled whenever applicable (34%, where A = 1% and F = 33%).

There is a close relation between urgency and type of arrival. While a part of urgent and the majority of semi-urgent patients can reach ED by foot, emergent patients are usually transported by ambulance. Initially, the walk-in and immobile patients are handled separately. Ambulance entrance area (crash room) is equipped with up-to-date life-saving facilities and resources, including first aid and resuscitation tools, instant trauma room and acute beds. Once it is possible, emergent patients are registered and thoroughly assessed. The most critical ones are just stabilized and immediately passed to the intensive care centre outside ED. Remaining urgent inpatients are transferred to the area of shared services. This main part of ED, available for all types of inpatients, provides specialized medical, diagnostic and laboratory services, as well as plastering and temporary confinement in non-acute, so called expectation beds with optional infusion therapy. Walk-in patients are registered and assessed in the main entrance area, where they wait for medical exam by particular specialist. Then they can either join the network of shared ED services or leave this unit. Approximately one fourth of inpatients is admitted to hospital, the rest is discharged. Consequently, mission of ED is composed from the following parts:

- Safety, i.e. continuous provision of acute care in all required areas (availability of service),
- Satisfaction, i.e. systematic application of standardized principles of quality control,
- Productivity, i.e. simultaneous effectiveness and efficiency of operations.

[1] College of Polytechnics Jihlava, Tolsteho 16, 586 01 Jihlava, Czech Republic, [jan.voracek, hana.vojackova, martina.kuncova]@vspj.cz .

[2] Hospital Jihlava, Vrchlickeho 59, 58633 Jihlava, Czech Republic, zazimald@nemji.cz .

Resultant global performance or utility function of ED must deeply balance these sub-goals. Investigation of related details and formulation of viable solutions are subjects of research interest for decades. The reason why structural and behavioural aspects of ED operations are so intensively analysed, discussed and improved is that the throughput and quality of work in this unit considerably influence many processes in the rest of the hospital. The most frequently reported ED bottlenecks are irregularly, but continuously overcrowded waiting rooms, overloaded human resources and over-utilized facilities and equipment. The presence of patients, standing in corridors, permanently occupied ED beds, waiting rooms without empty chair or urgent patients, waiting for medical examination for more than one hour are their most striking measurable examples [1-5]. There are also evident qualitative consequences of sudden ED malfunctions. Long-time waiting patients have tendency to leave the unit without being seen by doctor, restless working staff is biased to premature decisions and early discharged patients must usually be readmitted shortly [6-8]. Although the structure of ED waiting time or utilization of its resources has been analysed and discussed for a long time, there is just a little visible progress over this period. Despite of continuous development in medicine and technology, mean waiting times in EDs remain constant or even become slightly longer [9]. Many research studies, e.g. [10-13] identified reasons of overcrowding and classified them with respect to the hospital boundary. External factors arise mainly from political and social reasons, like unequal access to primary care physicians, insurance limitations, aging population or subjectively motivated self-referrals. Internal reasons include, e.g., staff shortages, missing expertise, low hospital capacity, delayed provision of shared services or imperfect medical history of incoming patients. We believe that all these shortcomings cannot be solved purely on the level of core, i.e. clinical operations, but must be delegated to the hospital top management. Only this body can guarantee appropriate interoperability of administratively standalone units and take full advantage of economies of scale. Moreover, standardized, systematically planned and holistically evaluated organization-wide non-clinical processes, including especially quality management and continuous improvement, can efficiently facilitate collaboration among hospital and its external partners. There are several advanced managerial methodologies, possibly applicable on ED processes, such as Lean, Six Sigma, Theory of Constraints (TOC), Total Quality Management or ISO standardization [14, 15]. In this research, we adopted TOC-driven direct parametric optimization, followed by Lean, structural optimization and Six Sigma respectively. Briefly speaking, TOC maintains and gradually minimizes the most limiting process constraints ("*a chain is no stronger than its weakest link*"), which has a positive impact on overall throughput. Lean continuously adds value for customer by cutting any admissible non-value adding procedures (waste) like transportation, storage or unnecessary bureaucracy. Six Sigma strives to keep variation of selected key performance indicators within the range of six standard deviations between their means and nearest specification limits, which guarantee close-to-zero defects operations, i.e. the best achievable quality. Focused joint application of Lean (less steps) and Six Sigma (less defects) implicates simultaneous improvements of speed, quality and costs.

2 Characteristics of ED operations

Our ED model merges sample data, knowledge and process layouts from several regional hospitals. In the main area of shared services we placed the following medical exam rooms:

- Surgery, cardiology and internal medicine operating continuously,
- Traumatology, operating only in the morning shift,
- Urology and neurology, operating in the afternoon and night shifts.

If required, patients can be diagnosed by X-Ray, ultrasound or computerized tomography (CT). Necessary biological samples can be examined in external laboratories. Described structure of generic ED can be split to the set of roughly chronologically ordered activities, listed in table 1. Arrival rate of inpatients was derived from three months hourly sample data averaged on weekly basis, which introduces characteristics morning peaks with daily falling trend and weekly maxima on Mondays. Beyond this basic data, full specification of ED model incorporates also particular layouts and capacities (explicit distances, amounts of people, equipment or facilities) as well as conditional probabilities of atomic activities in a form $P(activity|severity)$. In addition to stationary values, we were particularly interested in the transient behaviour of ED processes, generating also undesirable overcrowding.

3 Methodology

Complexity of ED modelling is determined mainly by the following aspects:

- Different priorities of patients and interrelated course of their diagnostics,
- Multidisciplinary scope of ED,
- Heavily shared (i) facilities (rooms, beds, chairs), (ii) specialized technical infrastructure (diagnostics and medical equipment, IT), and (iii) human resources.

Activity	Duration[1] [*min*]	Resources
Medical assessment (triage) of walk-in patients	(1, 4, 10)	Nurses
Registration of walk-in patients	(4, 7, 15)	Nurse
Registration and urgent treatment in crash room	(6, 10, 20)	Doctors, Nurses,
Preparation of patient in specialized exam room	(3, 10, 15)	Nurses
Medical assessment in specialized exam room	(2, 8, 15)	Doctors
Medical assessment by external specialists (optional)	(5, 10, 30)	Specialists
Specification of the next treatment	(2, 8, 15)	Doctors
Complete cycle of laboratory investigations (optional)	(10, 15, 350)	*External*
X Ray (optional)	(15, 20, 30)	Technicians
CT scan (optional)	(2, 8, 15)	Technicians
Ultrasound (optional)	(20, 35, 60)	Technicians
Plastering (optional)	(10, 15, 25)	Technicians
Review of laboratory and diagnostics results	(3, 5, 10)	Doctors
Discharge or admit decision	(3, 5, 10)	Doctors
Discharge or admit administration	(3, 5, 10)	Nurses

[1]triangular distribution

Table 1 List approximately chronologically ordered activities in emergency department

Using the terminology of operations research, ED is multiserver, multiphase and multistage dynamic problem, which can be hardly solved analytically. Our final computational model was designed in accordance with the four-steps problem formalization methodology, proposed, e.g., by Mitroff in [17]. Real-world problem is initially (i) verbally characterized and quantified, then (ii) conceptualized, (iii) computationally modelled and validated and (iv) practically implemented and organizationally evaluated with respect to requirements form phase (i). Depending on the collected conclusions, subsequent similar rounds of modelling and simulation can be formulated and realized. One such cycle is composed from the following items:

- Problem identification, which comprises apposite description of existing situation, altogether with list of related limitations and quantified specification of expected structural or parametric improvements.
- Conceptual modelling is a collection of all types of problem-characterizing artefacts, from well-structured data and reports, through approximate, subjective and uncertain information to unstructured chunks of expert knowledge. These heterogeneous modelling elements are identified through facilitated meetings, mind mapping, terminological unification and standardization. Winning components are integrated to a set system diagrams, shared among stakeholders.
- Computational modelling realizes conversion of the most relevant subset of conceptual elements to concrete programming platform. Applied DE framework conveniently extends possibilities and expressive power of traditional statistically or optimization oriented techniques, belonging to management science [17]. This technology offers deeper insight to complex, interrelated industrial processes and allows their dynamic visualization. DE modelling diagrams consist of predefined pathways (e.g. processes, workflows or layout-driven operations) structurally composed from library elements like servers, dispatchers, queues, delays, synchronization points, etc. Process conflicts can be solved either interactively by replicated change of parameters and restart or by grid-based numerical optimization, leading, e.g., to the minimal waiting time or equal utilization of resources. Presentation layer of DE software is usually graphically rich and realistically mimics investigated operations in natural layout. Further details can be found in [18]. The late modelling stages include verification (testing the erroneous functionality) and validation of initial assumptions (checking the level of compliance with users' goals). DE models can be easily validated through historical data, when computational outputs are compared with available datasets and related error statistics are enumerated. Alternatively, justified acceptance of behaviour by board of experts is also considered as a satisfactory way of validation.
- Model utilization stage analyses responses of well-tuned model from different managerial perspectives and suggests its viable enhancements. On top of one-shot interactive sessions, detailed experiments can employ also parametric sensitivity analyses, searching for parameters with the highest output impact. If external factors are of particular interest, their typical scenarios can be also designed and tested.

4 Design and results

We established and tested several cases, reflecting different levels of ED managerial maturity. In general, DE models are highly configurable - for example the presented ED implementation has 72 parameters. As it is impossible to work with all of them, one admissible strategy is to identify their key subset and run the desired experiments with varying single key parameter at a time. In the below outlined cases 4.1 – 4.3 we adjusted the numbers of the following resources:

- Doctors [N_D],
- Nurses [N_N],
- Specialists [N_S],
- Acute beds in crash room [N_{AB}],
- Expectation beds in area of shared ED services [N_{EB}].

Output functional and qualitative features were monitored through the set of following variables:

- Relative utilization of single resources, i.e. U_D, U_N, U_S, U_{AB} and U_{EB} in [$\%$],
- Total time spent in ED T_T in [min],
- Total waiting time T_W in [min],
- Length of the longest queue Q_L.

The queue length is an execution wide indicator, showing the maximal amount of walk-in patients, waiting at the same time either in registration or in assessment queue. If necessary, this number can be improved by adding more resources to the front administration. All outputs contain also qualitative information. For example, utilization of humans corresponds with level of their overloading, which negatively implies quality of work. Moreover, longer waiting times force some patients to leave ED without being seen by doctor, which usually shifts their problems to the future with potentially higher severity. Recalling the above-mentioned ED mission, it is evident, that any internal improvements cost something, but boosts productivity through higher efficiency and escalates quality with less defects and more satisfaction. Quantity and availability remain untouched, i.e. the patients obtain quality care by means of mature and efficient processes. Analysis and improvement of ED processes was realized according to the following cases 4.1–4.7 and obtained results are summarized in table 2.

4.1 Maximization of resources

This is an example of artificially overestimated numbers of all adjustable parameters, producing the best achievable performance. It is evident, that such configuration is practically unfeasible because of its high costs. On the other hand, this results can serve as appropriate benchmarking platform for the remaining cases.

4.2 Minimization of resources

This example represents a counterpart to case 4.1, i.e. forms the worst achievable, but still stable and conditionally applicable configuration. We can notice that the total net processing times ($T_T - T_W$) of the both initial cases are roughly comparable (97 vs. 118), but patients must wait more than three time longer because of inadequate capacities. High level of staff overloading is evident prerequisite for erroneous processes. Anyway, the minimal configuration is experimentally useful as it gradually discovers process bottlenecks and gives managers chance to find strategies for their removal. This option corresponds with the main mission of TOC methodology.

4.3 Realistic adjustment of resources

This analysis shows satisfactory parameterization of ED model. Although workload of nurses and utilization of expectations beds are on the edge, we adopted this basic configuration for all remaining cases, because of assumed further improvements.

4.4 Lean

This case introduces Lean principles to the previous configuration. Although there are several non-value adding candidates for possible removal, like unnecessary or slow transportation, obsolete technology or too complicated processes, we concentrated only on purely administrative procedures from table 1. Our proposed changes are summarized in table 3. In accordance with Lean principles, we did not influence any clinical or diagnostic operations, because all of them can potentially add value to patients. From the managerial point of view, simulated results represent the final steady state, which is achieved only after relatively long and costly implementation period. On the other hand, overall positive impact of out simple Lean institutionalization is evident as all performance indicators have improved and some of them are now even comparable with ideal case 4.1.

4.5 Structural changes

In this experiment we suggest inclusion of the following new processes to the previously proposed adjustments:

- Fast care track, where a doctor and nurse are dedicated entirely for uncomplicated, non-urgent patients,
- Introduction of mobile diagnostic devices, particularly X-Rays and ultrasounds,

Resultant higher throughput, faster treatment and shorter waiting times justify viability of such extensions.

Scenario		Resources, numbers and relative utilizations										Outputs		
		Doctors		Nurses		Specialists		Acute beds		Non-acute beds		T_T	T_W	Q_L
		N_D	U_D	N_N	U_N	N_S	U_S	N_{AB}	U_{AB}	N_{EB}	U_{EB}			
4.1	Maximal	6	40	6	63	3	40	6	17	10	37	140	43	2
4.2	Minimal	2	86	4	83	1	86	2	53	6	86	260	142	5
4.3	Realistic	4	59	5	77	2	57	3	35	6	71	160	57	3
4.4	Lean	4	52	5	65	2	47	3	33	6	53	141	46	3
4.5	Structural	4	56	5	64	2	50	3	35	6	55	132	42	3
4.6	Six Sigma	4	57	5	66	2	48	3	34	6	51	132	42	3
4.7	Costs min.	4	55	5	64	2	50	2	43	5	68	143	53	3

Table 2 Settings and results of all proposed experimental cases. Utilization of resources is in [%]

4.6 Six Sigma

This is a refinement of Case 4.5, when Lean-optimised administrative operations are also temporally balanced. Although we get only negligible improvements, inclusion of Six Sigma is generally useful, as it increases process capability and repeatability.

Activity	Duration[1] [*min*]		
	Original	Lean	Lean Six Sigma
Medical assessment (triage) of walk-in patients	(1, 4, 10)	(1, 4, 8)	(1, 4, 7)
Registration of walk-in patients	(4, 7, 15)	(4, 7, 12)	(4, 7, 10)
Preparation of patient in specialized exam room	(3, 10, 15)	(3, 8, 13)	(3, 8, 12)
Discharge or admit administration	(3, 5, 10)	(3, 5, 8)	(3, 5, 7)

[1]triangular distribution

Table 3 Selected adjustable administrative activities

4.7 Costs minimization

The last experiment evaluates economic effects of reduction of the both types of beds. Excellent performance from the previous case is spoiled in exchange for certain savings. This last alternative demonstrates possible practical closing of set of simulated analyses, when model, previously semi-optimized from performance viewpoints, is finally reconsidered also financially.

The total length of simulation was one month with one minute time step, i.e. we processed 43 200 samples for each experimental case. Experiments were implemented in DE simulation and visualization system SIMIO [19] under our institutional educational grant. This software has also excellent presentation possibilities, which make the outputs more comprehensible and attractive. Sample screenshots are in figure 1.

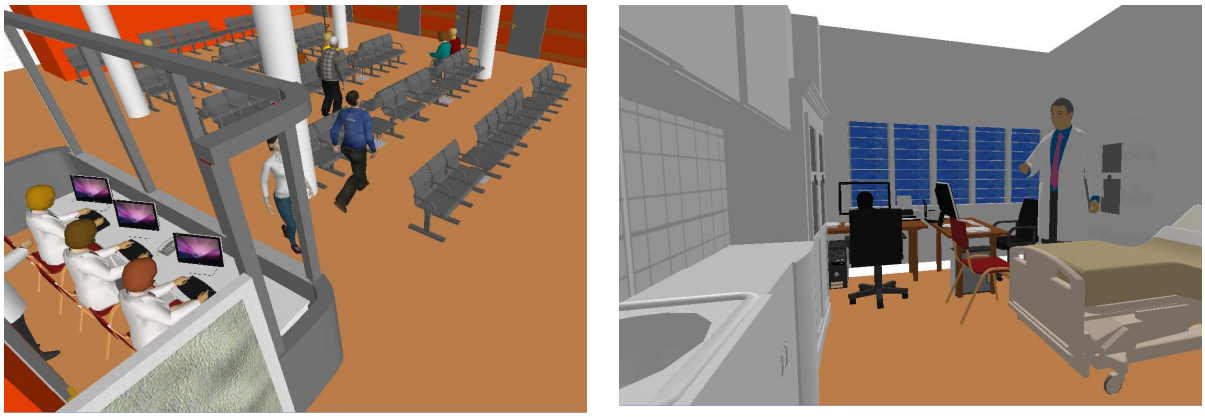


Figure 1 Examples of 3D visualization layer of emergency department model

5 Conclusions

We proposed an innovative way of clinical processes adjustment in hospital emergency department. Instead of more traditional qualitative, approximate or steady state analyses, our approach is based on modelling and interactive computational simulations of investigated phenomena. Using such technique, beyond the stationary behaviour, we could examine also related transient characteristics, which are essential in life-critical domains. We also presented several conceptual examples, outlining inclusion of selected standardized managerial frameworks, which are still rarely utilized in this sector. Straightforward and well justified inclusion of these tools to everyday practice supports efficient collaboration of hospitals with their supply chain partners, insurance companies and public health policy planning institutions. We believe that such multidisciplinary contribution with high application potential could address both healthcare scientists and practitioners. Our future plans include DE modelling of relations among different hospital departments with respect to the optimization of globally shared resources.

Acknowledgements

This research was supported by the College of Polytechnics, Jihlava, Czech grant "*Knowledge-based modelling and its applications*".

References

[1] Derlet, R. W, Richards, J. R and Kravitz, R. L.: Frequent crowding in US emergency departments. *Acad Emerg Med*. 2001; **8**(2001), 151–155.

[2] McCarthy, M. L, Aronsky, D., Jones I, Miner. J. R., Band, R. A., Baren, J. M., Desmond, J. S., Baumlin, K. M., Ding, R and Shesser, R.: The emergency department occupancy rate: a simple measure of emergency department crowding. *Ann Emerg Med*. **51** (2008), 15-24.

[3] Pines, J. M., Hilton, J. A, Weber E. J., et al.: International perspectives on emergency department crowding. *Academic emergency medicine* **18** (2011), 1358-70.

[4] Richardson, D. B. and Mountain, D.: Myths versus facts in emergency department overcrowding and hospital access block. *Med J Aust* **190** (2009), 369-374.

[5] Welch S and Davidson S.: Exploring new intake models for the emergency department. *Am J Med Qual* **25**(2010), 172-180.

[6] Ackroyd-Stolarz S, Read Guernsey J, Mackinnon NJ, et al.:The association between a prolonged stay in the emergency department and adverse events in older patients admitted to hospital: a retrospective cohort study. *BMJ quality & safety* 2011;20 (2011), 564-569.

[7] Cameron, P. A., Schull,M. J. and Cooke, M. W.: A framework for measuring quality in the emergency department. *Emerg Med J* doi:10.1136/emj.2011.

[8] Gilboy N., Travers D. A. and Wuerz R. C.: (1999). Re evaluating triage in the new millennium: A comprehensive look at the need for standardization and quality. *Journal of Emergency Nursing* **25**(1999), 468-473.

[9] Hing, E. and Bhuiya, F.: Wait Time for Treatment in Hospital Emergency Departments: 2009. *NCHS Data Brief, U.S. Department of Health & Human Services* **102** (2012).

[10] Kamali, M. F., Jain, M, Jain, A. R. and Schneider, S. M. Emergency department waiting room: many requests, many insured and many primary care physician referrals. *Int J Emerg Med*. **6**(2013):35.

[11] Jayaprakash, N, O'Sullivan, R, Bey, T, Ahmed, S. S. and Lotfipour, S., Crowding and delivery of healthcare in emergency departments: the European perspective. *West J Emerg Med*. **10** (2009), 233-239.

[12] Rose, L., Gray, S., Burns, K., Atzema, C., Kiss, A., Worster, A., Scales, D. C., Rubenfeld, G and Lee, J. Emergency department length of stay for patients requiring mechanical ventilation: a prospective observational study. *Scand J Trauma Resusc Emerg Med*. **20** (2012):30.

[13] Stead, L. G., Jain, A. and Decker, W. W.: Emergency department over-crowding: a global perspective. *Int J Emerg Med* **2** (2009), 133-134.

[14] Dahlgaard, J. J. and Dahlgaard-Park, S. M.: Lean production, six sigma quality, TQM and company culture. *TQM Magazine* **18** (2006), 263-281.

[15] Gareth, J. and George, J.: *Contemporary Management, 7th edition*. McGraw-Hill/Irwin; 2010.

[16] Mitroff I.I. and Kilmann, R.H.: *Methodological Approaches to Social Science: Integrating Divergent Concepts and Theories*. Jossey-Bass, San Francisco, 1978.

[17] Anderson, D.R., Williams, T. A., Sweeney, D. J., Martin, R. K.: *An Introduction to Management Science: A Quantitative Approach to Decision Making (12th Edition).* South-Western College Pub, 2007.

[18] Banks, J., Carson, J.S., Nelson, J. L. and Nicol, D. M.: *Discrete-Event System Simulation, 5/E,* Prentice Hall, 2010.

[19] Kelton, W. D., Smith, J. S., Sturrock, D. T. and Verbraeck, A. *Simio and Simulation: Modelling, Analysis, Applications*. McGraw-Hill, 2011.

Mean-Variance Distance Based Stock Market Networks in Portfolio Optimization

Tomáš Výrost[1], Štefan Lyócsa[2]

Abstract. Market networks have been used to describe relationships among market returns of assets. Although the economic relevance of market networks has been showed in many studies (e.g. Mantegna, 1999; Onnela et al., 2004, Lyócsa et al., 2012) applications had been rare. One of few exceptions is the study of Onnela et al. (2003) who showed that the set of equities selected by the Markowitz's mean-variance portfolio optimization technique can be related to the vertices on stock market networks, i.e. distant vertices (path length) are selected in the Markowitz's mean-variance portfolio as well.
We propose a Black and Litterman (1992) type portfolio optimization model, where return estimates are derived from information given by the stock market network. Instead of usual correlations, the networks are constructed using extreme coexceedance and mean-variance distances between assets. The performance of our model is compared with standard Markowitz's mean-variance portfolio. This study has the potential to contribute to the existing studies on stock market networks, as it directly links stock market networks to portfolio optimization problems.

Keywords: Stock market networks, portfolio, emerging markets, Black-Litterman model.

JEL Classification: G15, D85, L14
AMS Classification: 62P20, 94C15, 91G70

1 Introduction

In accordance with the classical mean-variance portfolio theory [12], international portfolio diversification may offer stock investors better risk-return characteristics compared to the solely domestic asset allocation. Unfortunately, recent empirical evidence suggests an increase in market integration and as so, decrease in diversification benefits (e.g. [9]; [6]; [7]; [3]). As international market integration increases, one might assume that more attention will be given to the methods of portfolio diversification. Since the seminal work of Markowitz [12] countless methods for portfolio diversification were developed, which approach the known limitations of the Markowitz's model from different perspectives. In this preliminary study, we will limit ourselves to a short overview of existing alternative approaches to portfolio optimization.

The first group of authors proposes models, which change investor's objective function. For example, the objective function might optimize not only risk but also expected return. More complex models include optimization of conditional Value-at-Risk (e.g. [17]). Another group of studies considers different risk measures (e.g. downside risk) or the way how market correlations are estimated (e.g. [8]). The third group of studies aims at changing the original optimized model, with the goal to take into account alternative relevant information. Just to mention few relevant studies, Qian and Gorman [16] have proposed a model which allows investors to take into account not only their personal beliefs about the expected returns but also about the correlation and volatility of returns. Almgren and Chriss [1] have designed a less structured model where it is not required to know exact expected returns, but only the rank of returns. A more general approach has been devised in Meucci [13], where one can exploit information about the expected values of exogenous factors which in turn influence stock market returns. In the presented paper, we attempt to contribute to this strand of the literature by incorporating information from the topological properties of stock market networks into portfolio optimization. In this short paper we will present application where we will utilize the standard Black and Litterman [4] model.

The remainder of the paper is organized as follows. Sections 2 describes the proposed coexceedance and mean-variance distance based stock market networks and suggest a vertex characteristic which might be exploited in portfolio optimization. In Section 3 we briefly describe the standard Markowitz optimization problem [12] and the extension of Black and Litterman [4]. Section 4 presents the data used in the empirical part of the study

[1] University of Economics, Faculty of Business Economics, Department of Financial Management, Tajovského 13, 04001 Košice, tomas.vyrost@euke.sk.
[2] University of Economics, Faculty of Business Economics, Department of Quantitative Methods, Tajovského 13, 04001 Košice, stefan.lyocsa@euke.sk.

and the design of the simulation; here we also evaluate the results from our simulations. The final section concludes.

2 Stock market networks

Since the seminal work of Mantegna [11], relationships among stocks have been modelled using correlation based networks. Relationships among stock returns may be represented with a correlation matrix ($\rho_{ij} \in \mathbf{C}$), which is a complete graph, with vertices being stocks and edges weights derived from the following correlation based distance measure [11]: $d_{ij} = (2(1-\rho_{ij}))^{0.5}$. To filter out information from noisy correlations (and hence distances) a subgraph is usually extracted from the distance matrix **D**, e.g. the minimum spanning tree (MST).

The topological properties of the resulting graphs are exploited for further analysis. For example Onnela et al. [14] showed that the set of equities selected by the Markowitz's model can be related to the vertices on stock market networks (in MSTs), i.e. distant vertices (path length) are selected in the Markowitz's mean-variance portfolio as well.

Our goal was to construct networks, which will describe the structure of extreme negative co-movement between stocks, as during these times, the diversification is most needed. To construct such networks we used two ideas: (i) mean-variance distances measure, and (ii) the principle of extreme coexceedance. The normalized mean-variance return distances (rr_{ijt}, $i \neq j$) for each pair of stocks is calculated as in Eun and Lee [7]:

$$rd_{ijt} = \left| r_{it} - r_{jt} \right| \tag{1}$$

$$sd_{ijt} = \left| s_{it} - s_{jt} \right| \tag{2}$$

$$w(rd)_{ij} = \left(\sum_{t=1}^{T} rd_{ijt}^2 / \left(\sum_{t=1}^{T} rd_{ijt}^2 + \sum_{t=1}^{T} sd_{ijt}^2 \right) \right)^{0.5} \tag{3}$$

$$w(sd)_{ij} = \left(\sum_{t=1}^{T} sd_{ijt}^2 / \left(\sum_{t=1}^{T} rd_{ijt}^2 + \sum_{t=1}^{T} sd_{ijt}^2 \right) \right)^{0.5} \tag{4}$$

$$rr_{ijt} = \left(\left(rd_{ijt} / w(rd)_{ij} \right)^2 + \left(sd_{ijt} / w(sd)_{ij} \right)^2 \right)^{0.5} \tag{5}$$

where r_{it} is the return of i-th stock in weak t and s_{it} is the standard deviation of i-th stock's return in weak t.

A *negative coexceedance network* at time t can be defined as a graph $G(V, E_t)_t$, where V is the set of all stocks and E_t is a set of edges. An edge $E_t(i,j)$ between stock i and j exists, when $r_{it} < r_{it}(5) \wedge r_{jt} < r_{jt}(5)$, where $r_{it}(5)$ and $r_{jt}(5)$ are the 5^{th} percentile of returns for market i and j respectively. The principle of return coexceedance is simiar as in Bae et al. [2]. Next, we define a *mean-variance return distance weighted negative coexceedance network* (MV negative coexceedance network in short) by using the mean-variance return distances as edge weights in the negative coexceedance network.

Note, that the $G(V, E_t)_t$ can be described via an adjacency matrix and that index t represents a particular week. In the simulations which will follow, we will sum adjacency matrices over 52 weeks of t to form a new adjacency matrix $\mathbf{W}_t$, where $w_{ijt} \in W_t$ denotes weights, i.e. number of times extreme negative coexceedance occurred between stocks i and j during the 52 week period for the negative coexceedance network. For the MV negative coexceedance networks, each nonzero weight w_{ijt} is replaced by the sum of the risk-return distances over the 52 week period for the negative coexceedance network.

With regard to the portfolio optimization, it seems plausible to keep stocks, which do not have extreme coexceedances of returns with other stocks in the portfolio. In the negative coexceedance networks, this vertex property may correspond to low vertex centrality. Let $d(i,j)$ denote the shortest path between vertex i and j (unweighted or weighted by the sum of risk-return distances). The Boldi and Vigna [5] harmonic centrality measure seems an adequate candidate as it is also suitable for unconnected networks:

$$\sum_{d(i,j)<\infty, i \neq j} \frac{1}{d(i,j)} \tag{6}$$

3 Portfolio optimization

To explore our approach in a portfolio setting, we use three types of generally used portfolio models as a basis for comparison for our coexceedance and distance-based approach. Each of these approaches differs in the way the portfolio weights are assigned.

The simplest approach assigns equal weights to all stocks in the portfolio. Although this presents a rather naive construction, it achieves some diversification benefits because of the sheer number of stocks the investor holds. The second approach consists of the calculation of the so called tangential portfolio. To define this portfolio, we first define the return (r_p) and risk (σ_p^2) of a portfolio of $k \in \mathbb{N}$ stocks as follows

$$r_p = \boldsymbol{\alpha}^T \mathbf{r}, \qquad \sigma_p^2 = \boldsymbol{\alpha}^T \mathbf{C} \boldsymbol{\alpha} \tag{7}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^k$ is a vector of portfolio weights, $\mathbf{r} \in \mathbb{R}^k$ is a vector of stock returns and $\mathbf{C}$ is a $k \times k$ covariance matrix of individual returns. The portfolio risk is thus expressed as standard deviation (or variance) of portfolio returns. The tangential portfolio corresponds to the weights maximizing the Sharpe ratio, given by the amount of excess return per unit of risk

$$\frac{r_p - r_f}{\sigma_p} = \frac{\boldsymbol{\alpha}^T \mathbf{r} - r_f}{\sqrt{\boldsymbol{\alpha}^T \mathbf{C} \boldsymbol{\alpha}}} \tag{8}$$

where r_f is a risk free rate of return (usually the yield of short-term government debt).

The third approach uses a combined objective function maximizing

$$\boldsymbol{\alpha}^T \mathbf{r} - \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{C} \boldsymbol{\alpha} \tag{9}$$

As the elements of the covariance matrix will be quite small, and the risk-aversion factor is 0.5, this represents a rather aggressive choice seeking primarily higher returns.

The fourth alternative is based on our negative coexceedance stock market networks. This model is the most complex model tested, as the optimization is not done directly – instead, we employ a Black – Litterman correction of expected returns, which are then used as inputs in a Sharpe ratio maximizer.

Finally, the fifth model employs a Black – Litterman correction of expected returns on a MV negative coexceedance network.

The Black – Litterman algorithm has several steps: first, initial portfolio weights are selected, which may be based on factor model or individual investor preferences. These are then used in reverse optimization to calculate the expected return estimates. These expected returns are then combined via a Bayesian approach with prior historical market estimates of returns ($\boldsymbol{\pi} \in \mathbb{R}^k$) to obtain posterior estimates, used in the actual optimization. Thus, the whole purpose of the Black – Litterman algorithm is to obtain better approximation of real stock returns $\boldsymbol{\mu} \sim N(\boldsymbol{\pi}, \mathbf{C})$.

In our case, the initial portfolio weights ($\mathbf{w} \in \mathbb{R}^k$) have been selected proportionally to the difference of harmonic centralities of vertices in a stock market network from the maximum of these centralities. Thus, the vertices with high centralities will have low weights and vice versa, as the objective of efficient diversification should focus on non-central nodes. The initial weights have been then used in the reverse optimization to obtain the estimated returns

$$\mathbf{r}' = \frac{1}{2} \mathbf{C} \mathbf{w} \tag{10}$$

The investor views, expressing our expectations regarding the real stock returns $\boldsymbol{\mu}$, are then written using the pick matrix $\mathbf{P}$ (in our application equal to the identity matrix $\mathbf{I}_k$) as

$$\mathbf{P}\boldsymbol{\mu} \sim N\left(\mathbf{r}', \frac{1}{2} \mathbf{P}^T \mathbf{C} \mathbf{P} \right) \tag{11}$$

By using the Black – Litterman's results, we obtain the adjusted expected returns

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\pi} + \mathbf{C}\mathbf{P}^T\left(\mathbf{P}\mathbf{C}\mathbf{P}^T + \frac{1}{2}\mathbf{P}\mathbf{C}\mathbf{P}^T\right)^{-1}\left(\frac{1}{2}\mathbf{C}\mathbf{w} - \mathbf{P}\boldsymbol{\pi}\right) \tag{12}$$

which are then used in an optimization seeking weights yielding the maximum Sharpe ratio (tangential) portfolio. Hence, the calculation is similar to the second approach, but using stock market network based Black – Litterman adjustments to the expected returns.

4 Results

An empirical simulation study may allow us to observe and compare basic portfolio characteristics of our approach with that of a standard portfolio selection models. Our sample of weekly returns and standard deviations of daily returns for a given week (T = 415) consists of Central and Eastern European equities from January 2nd, 2006 to December 31st, 2013. A total of 125 equities were selected from six CEE and a German stock market; Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovenia, and Germany (constituents of the DAX index as of December 2013). A complete list of equities is available upon request. Daily closing prices are obtained from Datastream.

Our simulation proceeds as follows:

1. We randomly select N = 15 equities;
2. Using returns from the previous 52 weeks we construct a negative coexceedance network (Section 2) and calculate harmonic centralities for all vertices.
3. We estimate portfolio weights for all five approaches – the equally weighted, tangential, combined objective, and the two Black – Litterman adjusted weights (Section 3).
4. We evaluate the portfolio characteristics among the portfolios one week ahead (ex ante).
5. We repeat steps 1 – 4 using next 52 weeks of observations (rolling window is drifted one weak ahead).

The whole process has been repeated 100 times for the whole sample, to randomize the choice of stocks each week. The results of our simulation study are shown in Table 1 and 2. In Panel A of Table 1 it is straightforward to see that the combined strategy is very aggressive as most of the time only one or two stocks are selected into the portfolio. On the other hand, the Sharpe and coexceedance portfolios select more stocks, which provide investors more protection from unsystematic risks. In Panel B of Table 1 we compare standard portfolio characteristics. Note that although the coexceedance strategy provides larger risk compared to the equally weighted portfolios and portfolio according to Sharpe, it also provides investor with much higher cumulative returns.

	Min	1Q	Median	Mean	3Q	Max	SD	Skew.	Kurt.	CR	CCR
Panel A: Number of selected stocks											
Sharpe	3.90	7.01	8.72	8.59	10.00	13.72					
Combined	1.00	1.00	1.00	1.56	2.00	4.54					
Coexceedance	1.00	3.00	5.00	5.04	7.00	11.38					
Coex. weighted	1.00	3.00	5.00	4.99	6.99	11.33					
Panel B: Returns – one week ahead											
Equally	-0.203	-0.012	0.003	0.000	0.015	0.100	0.029	-1.354	8.663	0.073	-0.061
Sharpe	-0.167	-0.011	0.001	-0.001	0.012	0.093	0.025	-1.248	8.731	-0.290	-0.297
Combined	-0.226	-0.024	0.000	0.001	0.026	0.283	0.054	0.367	5.613	0.461	0.612
Coexceedance	-0.253	-0.014	0.002	0.001	0.018	0.331	0.045	0.754	20.76	0.433	0.358
Coex. weighted	-0.257	-0.014	0.002	0.001	0.018	0.334	0.046	0.721	20.02	0.438	0.331

Table 1 Descriptive statistics of out-of-sample portfolios

Note: all results are averaged across 100 iterations. *CR* – denotes cumulative portfolio returns at the end of the sample. *CCR* – denotes cumulative compounded portfolio returns at the end of the sample.

	$AC(ret)$	$AC(ret^2)$	Pearon's correlation				Spearman's correlation			
			Sh.	Com.	Coex.	Co. w.	Sh.	Com.	Ex.	Co. w.
Equally	0.10	0.19	0.94	0.74	0.85	0.84	0.86	0.69	0.80	0.80
Sharpe	0.21	0.17		0.73	0.82	0.82		0.62	0.73	0.73
Combined	0.11	0.37			0.88	0.88			0.85	0.85
Coexceedance	0.08	0.46				0.99				0.99
Coex. weighted	0.09	0.45								

Table 2 Time series characteristics and correlation of out-of-sample portfolio returns

Note: series are derived as averages across 100 iterations. $AC(ret)$ – denotes first order autocorrelation coefficient of returns, $AC(ret^2)$ – denotes first order autocorrelation coefficient of squared returns.

Table 2 offers some standard time series characteristics. Autocorrelation of returns and squared returns is positive, which is typical for financial time series. Finally, it might be interesting to observe, whether the returns between different strategies are correlated. From the reported results (see Table 2) it seems obvious that the returns from the combined strategy are less correlated with other strategies (see Figure 1).

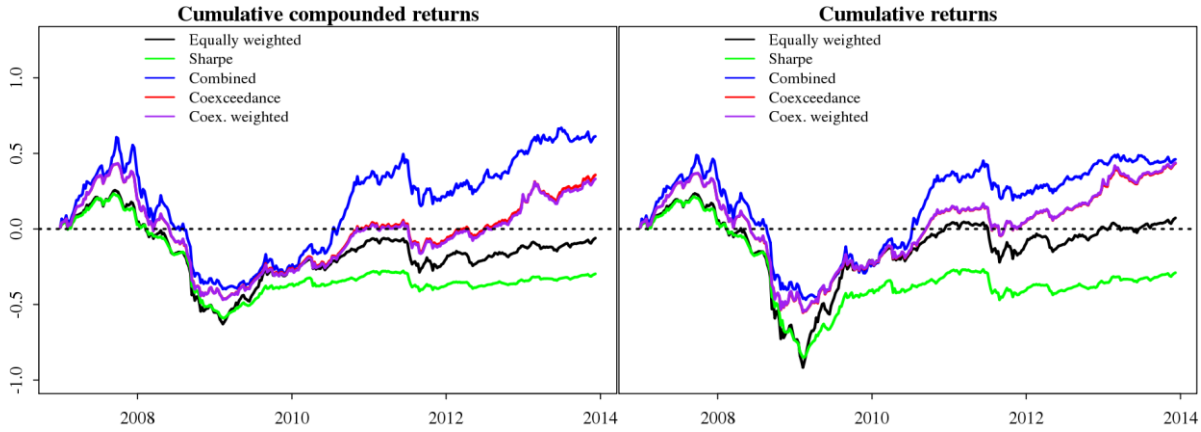


Figure 1 Portfolio performance

5 Conclusion

Our paper primarily seeks to search for a practical application of stock market networks modelling by incorporating the notions of centrality into the portfolio selection problem. Although the literature describing stock market networks is numerous, specific practical applications are rare.

We compared the proposed approach with three more conventional alternatives – an equally weighted portfolio, a portfolio based on the maximization of the Sharpe ratio (thus seeking the best reward/risk ratio) and a portfolio maximizing a combined objective, allowing for the substitution between risk and returns.

The incorporation of network properties into the portfolio setting has been conducted using the Black and Litterman [4] approach, which adjusts the historical expected returns by quantitative investor views. By assuming a negative relationship between risk and return, we use the harmonic centrality in a network to reverse-optimize expected returns. These are then used to form posterior return estimates, entering portfolio optimization.

The results of our analysis show, that despite not having the highest average cumulative (compounded) returns, our approach leads to substantially higher values than the equally weighted and tangential portfolios. Especially the case of tangential (maximum Sharpe ratio) portfolios is interesting, as it is one of the most common choices in portfolio theory.

In terms of averages, our portfolios come second, after the combined objective portfolios. However, these were obtained from a model that is rather aggressive, focusing primarily on expected returns. As was shown in the previous section, the combined objective portfolios have highest risk measured by standard deviation.

Interestingly, the results also suggest very minor differences between unweighted negative coexceedence networks and MV negative coexceedence networks. The edge weights based on the mean-variance distances do not provide much additional information, as the structure of the network based on coexceedance alone does equally well in portfolio applications.

The real practical use of the coexceedance and mean-variance distance based stock market networks in portfolio optimization as shown in this paper is not completely straightforward. First, the simulation study could be extended with respect to the number of iterations in order to get better statistical properties of the results. Second, the effect of portfolio size has not been included in this study – this also has the potential to influence the results. Third, this paper does not deal with the optimal choice of parameters both in the combine objective (we used the constant of 1/2 to adjust for risk) and Black and Litterman model [4]. Particularly the Black and Litterman model [4] has several constants that should be set according to the investor subjective preference, which is in general difficult to assess in a simulation study.

Despite the existence of these potential extensions, our paper provides empirical evidence for practical application of graphs in portfolio optimization – the structure captured in negative coexceedence networks seems to provide relevant information that might be practically exploitable.

Acknowledgements

The authors acknowledge the funding support of the Slovak Research and Development Agency (APVV Project No. APVV-0666-11).

References

[1] Almgren, R., and Chriss, N.: Optimal execution of portfolio transactions. *Journal of Risk* **3**, 2 (2006), 5-39.
[2] Bae, KH., and Karolyi, AG., and Stulz, RM.: A New Approach to Measuring Financial Contagion. *The Review of Financial Studies* **16**, 3 (2003), 717-763.
[3] Baumöhl, E, and Lyócsa, Š.: Volatility and dynamic conditional correlations of worldwide emerging and frontier markets. *Economic Modelling* **38** (2014): 175-183.
[4] Black, F., and Litterman, R.: Global portfolio optimization. *Financial Analysts Journal* (1992): 28-43.
[5] Boldi, P., and Vigna, S.: In-Core Computation of Geometric Centralities with HyperBall: A Hundred Billion Nodes and Beyond, arXiv:1308.2144.
[6] Chelley-Steeley, P.: Equity market integration in the Asia-Pacific region: A smooth transition analysis. *International Review of Financial Analysis* **13**, 5 (2004): 621-632.
[7] Eun, Ch.S. and Lee, J.: Mean–variance convergence around the world. *Journal of Banking & Finance* **34**, 4 (2010): 856-570.
[8] Ledoit, O., and Wolf, M.: Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management* **30**, 4 (2004): 110-119.
[9] Longin, F, and Solnik, B.: Is the correlation in international equity returns constant: 1960–1990?. *Journal of international money and finance* **14**, 1 (1995): 3-26.
[10] Lyócsa Š., and Výrost T., and Baumöhl E.: Stock market networks: The dynamic conditional correlation approach. *Physica A: Statistical Mechanics and its Applications* **391**, 16 (2012), 4147-4158.
[11] Mantegna RN.: Hierarchical structure in financial markets. *European Physical Journal B* **11**, 1 (1999), 193–197.
[12] Markowitz, H.: Portfolio selection*. *The journal of finance* **7**, 1 (1952): 77-91.
[13] Meucci, A.: Enhancing the Black–Litterman and related approaches: Views and stress-test on risk factors. *Journal of Asset Management* **10**, 2 (2009): 89-96.
[14] Onnela JP., and Chakraborti A., and Kaski K., and Kertész J.: Dynamic asset trees and Black Monday. *Physica A: Statistical Mechanics and its Applications* **324**, 1-2 (2003), 247-252.
[15] Onnela JP., and Kaski, K., and Kertész J.: Clustering and information in correlation based financial networks *European Physical Journal B* **38**, 2 (2004), 353–362.
[16] Qian, E., and Gorman, S.: Conditional distribution in portfolio theory. *Financial Analysts Journal* (2001): 44-51.
[17] Rockafellar, R. T., and Uryasev, S.: Optimization of conditional value-at-risk. *Journal of risk* **2** (2000): 21-42.

The conditional dependence structure among precious metals: a copula-GARCH approach

Stanisław Wanat[1], Monika Papież[2], Sławomir Śmiech[3]

Abstract. The aim of the paper is to analyse the conditional dependence structure among precious metal returns using a copula-DCC-GARCH approach. Conditional correlation matrices are used to identify the states of the precious metals market by assuming that a given state of the market corresponds to a typical pattern of the conditional dependence structure. Cluster analysis allows for pointing at transition points among the market states, that is the points of drastic change in the conditional dependence structure. The application of the methodology described above to the period between 1997 and 2013 indicates three market states of four major precious metals (gold, silver, platinum and palladium). The results obtained reveal a sudden increase in dependencies among precious metals at the turn of April and May 2004.

Keywords: precious metals, dependence structure, copula-GARCH, market states.

JEL Classification: C58, Q02
AMS Classification: 91G70, 62P20

1 Introduction

The existing research on precious metals focuses mainly on gold and silver. The analysis of the prices of precious metals can be divided into two areas. The first one covers the analysis of the relationship among prices of precious metals. Ciner [2] finds evidence of the disappearance of the long term relationship between gold and silver in the 1990s. Their conclusion is contested by Lucey and Tully [10], who say that this relationship strengthens and weakens over time but is prevalent in the long run. Similarly, Sari et al. [20] notice a strong relationship between gold and silver. Śmiech and Papież [25] show that causality among the prices of gold, silver, platinum and copper change in the period 2000-2011. Also Papież and Śmiech [14] examine causality in mean and variance between commodity prices (including metal prices) and financial market prices.

The second area of analysis is connected with examining the volatility of returns of the precious metals. Hammoudeh and Yuan [5] show that gold and silver have similar volatility persistence globally, but there is no leverage effect in gold and silver prices. Sari et al. [21] examine the co-movements and information transmission among the spot prices of four precious metals (gold, silver, platinum, and palladium), the oil price, and the US dollar/euro exchange rate. Hammoudeh et al. [6] examine the conditional volatility and correlation dependence for four major precious metals, and they find that almost all of them are weakly responsive to news spilled over from other metals in the short run. Morales and Andreosso-O'Callaghan [11] find that in terms of volatility spillovers, an asymmetric effect is observed; gold tends to dominate the markets and the evidence favouring the case of other precious metals influencing the gold market is weak. Cochran et al. [3] show that events taking place during the post-September 2008 period increased the volatility in gold, platinum, and silver returns. Sensoy [22] claims that the turbulent year of 2008 had no significant effect on the volatility levels of gold and silver, although caused an upward shift in the volatility levels of palladium and platinum. Using the consistent dynamic conditional correlations, he shows that precious metals became strongly correlated with each other during the last decade, which reduces the diversification benefits across them and indicates a convergence to a single asset class.

The objective of this study to identify the states of the precious metals market (gold, silver, platinum and palladium) and to present their temporal evolution in the period from September 22, 1997 to February 13, 2014. Since our sample period covers the recent global financial crisis, we want to examine whether the market states are affected by the financial crisis. The process of identifying the states of the precious metals market and analysing their temporal evolution is based on the conditional dependence structure using a copula-DCC-GARCH methodology.

This allows us to address several questions, which might be of interest to both investors and researchers:

- Is the dependence among prices in the precious metals markets stable or does it undergo changes?

[1] Cracow University of Economics, 27 Rakowicka St., 31-510 Cracow, Poland e-mail: wanats@uek.krakow.pl.
[2] Cracow University of Economics, 27 Rakowicka St., 31-510 Cracow, Poland e-mail: papiezm @uek.krakow.pl.
[3] Cracow University of Economics, 27 Rakowicka St., 31-510 Cracow, Poland e-mail: smiechs@uek.krakow.pl.

- Are the changes in relations among precious metals prices evolutionary or drastic?
- What are the causes of drastic changes in relations among these prices?

The paper contributes to the existing literature in the following aspects.

Firstly, most analyses of the precious metals market conducted so far are based on standard multivariate GARCH (MGARCH) models (see e.g. Hammoudeh et al. [6], Morales and Andreosso-O'Callaghan [11], Sensoy [22], Silvennoinen and Thorp [24]), which assume that standardized innovations follow a multivariate elliptical distribution. In case of the multivariate normal distribution all marginal distributions must be normal, and multivariate Student's t distribution imposes, also often unrealistically, the same degrees of freedom for all marginal distributions.

However, a copula-based multivariate GARCH model used in this study allows for modelling the conditional dependence structure when standardized innovations are non-elliptically distributed. Thus, it makes it possible to model the volatility of particular metals using univariate GARCH models with different standarized residual distribution. Generally, copulas allow the researcher to specify the models for the marginal distributions separately from the dependence structure that links these distributions to form a joint distribution. They offer a greater flexibility in modelling and estimating margins than while using parametric multivariate distributions (see e.g. Nelsen [13], Joe [7]).

Secondly, at present a copula-GARCH methodology is widely used in the analysis of financial time series (see e.g. Aloui et al. [1], Lee and Long [8], Li and Yang [9], Patton [15], Philippas and Siriopoulos [17], Serban et al. [23], Wu et al. [26], Zolotko and Okhrin [27] and for a review Patton [16]). However, in most studies on the precious metals market, copula methodologies are used to analyse the dependencies among single metal markets and other markets (see e.g. Reboredo [18, 19]). This study is based on conditional correlations using a copula-GARCH methodology to investigate the dynamics of conditional dependence structure among precious metals. It also attempts to identify the states of the market on the basis of these conditional correlations and to follow their temporal evolution. To the best of our knowledge, such approach has not been applied to investigate dependencies in the precious metals markets so far.

The paper is organised as follows. Section 2 describes the data and econometric methodologies employed. Empirical results are discussed in Section 3, and the conclusions are presented in the last section.

2 Methodology

The dynamic relationship among precious metals is analysed with the use of a copula-DCC-GARCH model for daily log-returns. In this approach, multivariate joint distributions of the return vector $r_t = (r_{1,t},...,r_{k,t})'$, $t = 1,...,T$, conditional on the information set available at time $t-1$ (Ω_{t-1}) is modelled using conditional copulas introduced by Patton [15]. This model takes the following form:

$$r_{1,t} \mid \Omega_{t-1} \sim F_{1,t}(\cdot \mid \Omega_{t-1}),...,r_{k,t} \mid \Omega_{t-1} \sim F_{k,t}(\cdot \mid \Omega_{t-1}), \tag{1}$$

$$r_t \mid \Omega_{t-1} \sim F_t(\cdot \mid \Omega_{t-1}), \tag{2}$$

$$F_t(r_t \mid \Omega_{t-1}) = C_t\left(F_{1,t}(r_{1,t} \mid \Omega_{t-1}),...,F_{k,t}(r_{k,t} \mid \Omega_{t-1}) \mid \Omega_{t-1}\right), \tag{3}$$

where C_t denotes the copula, while F_t and $F_{i,t}$ respectively the joint cumulative distribution function and the cumulative distribution function of the marginal distributions at time *t*. Using a copula allows for separate modelling of marginal distributions and the dependence structure of vector r_t. In the empirical study elliptical copulas are used to describe the dynamics of the dependence structure, while conditional marginal distributions are modelled with the use of ARMA-GARCH models. Conditional correlation matrices R_t modelled with the use of DCC(1, 1) model (Engle [4]) are assumed to be the parameters of conditional copulas.

The states of the precious metals market are identified on the basis of conditional correlation matrices. It is assumed that a given market state corresponds to a typical pattern of the conditional dependence structure described by a conditional correlation matrix R_t. Transition points between market states, corresponding to drastic changes in the conditional dependence structure, are identified using Ward's method of cluster analysis and a similarity measure suggested by Münnix et al. [12], which allows us to quantify the difference of the correlation structure for two points in time.

3 Data and empirical results

The data used in this study consist of the daily (five working days per week) spot prices of gold (Gold), silver (Silv), platinum (Plat) and palladium (Pall) from September 22, 1997 to February 13, 2014. The source of data is Bloomberg, and all prices of precious metals are measured in US dollars per troy ounce. As usual, price return series are computed on a continuous compounding basis as $r_{i,t} = 100 \times (\log(P_{i,t} / P_{i,t-1}))$, where $P_{i,t}$ and $P_{i,t-1}$ are current and one-period lagged spot prices of precious metals. After eliminating the mismatching transaction days, we finally obtain 4185 log-returns for each series.

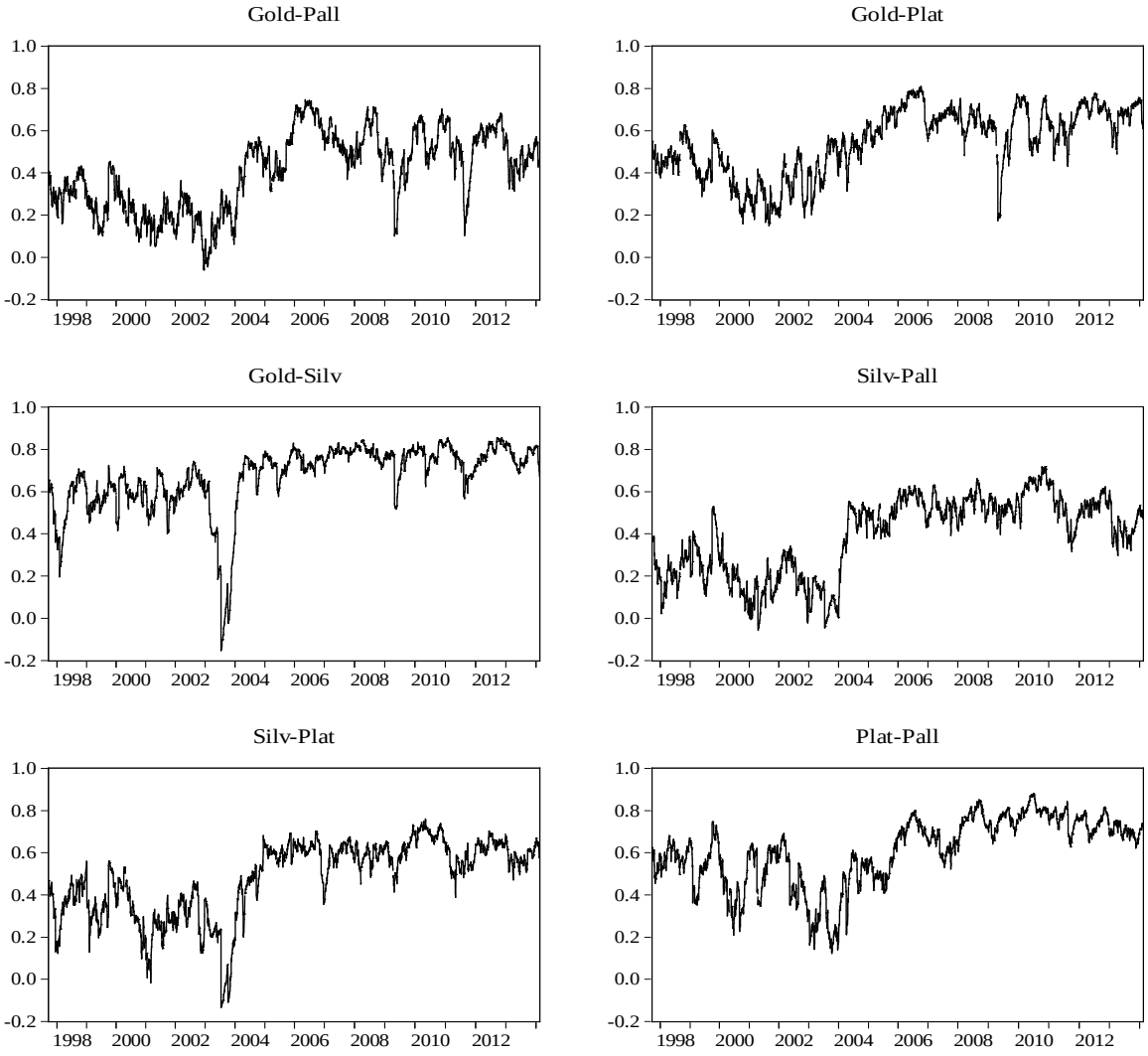


Figure 1 Dynamic correlations of precious metal returns

In the empirical study different variants of the AR-GARCH specification are considered for individual returns. Eventually, on the basis of information criteria, Student's t AR(1)-GARCH(1,1) model has been assumed for gold, normal AR(1)-GARCH(1,1) model has been assumed for silver, skewed Student's t ARMA(1,1)-GARCH(1,1) model has been assumed for platinum, and Student's t AR(1)-GARCH(1,1) has been assumed for palladium. On the other hand, Gauss and Student's t copulas have been considered in the analysis of the dynamics of dependencies among the rates of return, and, again on the basis of information criteria[4], Student's t has been chosen.

Conditional correlation matrices obtained with the use of the estimated model are applied to analyse the precious metals market. Dynamic correlations in this marked are presented in Fig. 1. Figure 2 shows the temporal evolution of the market states in the period from September 22, 1997 to February 13, 2014, obtained as a result

[4] The parameters of the model are assessed using R package "rmgarch" (version 1.2-6), developed by Alexios Ghalanos. The results can be obtained from the author on request.

of clustering conditional correlation matrices with Ward's method of cluster analysis. The left panel illustrates the division into two clusters (Rousseeuw's Silhouette internal cluster quality index equals 0, 5512), while the right panel illustrates the division into three clusters (Rousseeuw's Silhouette internal cluster quality index equals 0,3169). Structural changes in the precious metals markets in the analysed period are identified with the use of a similarity measure of correlation matrices (Münnix et al. [12]) and presented in Fig. 3 (the left panel illustrates the similarity of matrices distant from each other by a multiple of one quarter, the right panel – by a multiple of one year). Light shading denotes similar conditional correlation matrices and dark shading denotes dissimilar ones. If we assume that a point on the diagonal designates ''now'', then the similarity to previous times from this point can be found on the vertical line below this point, or the horizontal line to the left of this point.

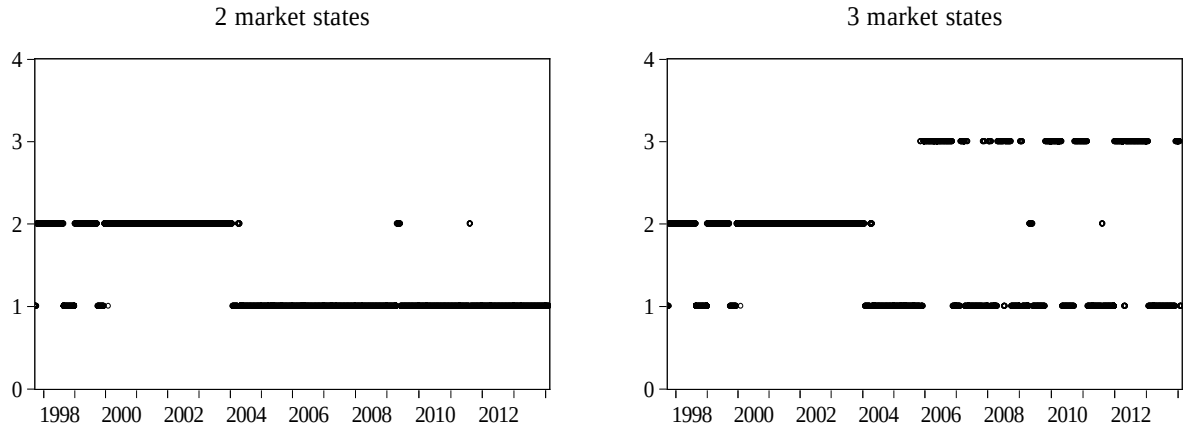


Figure 2 Temporal evolution of the market state

After analysing the dynamics of conditional correlation matrices (Fig. 1), it can be said that bilateral correlation among precious metals markets increased considerably in 2004 and has remained at this level since then. A similar conclusion can be drawn from the analysis of similarity maps showing conditional correlation matrices (Fig. 3). Particular areas in the matrix show how similar dependence measures in two periods are: the first one is on the horizontal axis and the second one on the vertical axis. Figure 3 presents the results for two options: the first option covers correlations calculated for sub-periods with 70 observations, while the second – for sub-periods with 250 observations. A low value of a similarity measure of correlation matrices (considerably darker shading) for 2004 together with its rise and maintaining this high level (lighter shading) indicate structural changes in precious metals markets in 2004. It confirms Sensoy's [22] hypothesis that precious metals will be a single asset class in the near future. This change in the precious metals markets in 2004 is also evident in clustering results (Fig. 2). Divisions into two and three clusters indicate a stable change of the market state in April 2004 (4/29/2004). For the last decade this market has not returned (with few exceptions) to the state from before 4/29/2004. The more detailed analysis of this decade (the division into 3 clusters) reveals two basis states with numerous transition points between them. On the basis of the results obtained, it can be concluded that the global financial crisis from 2008 has not considerably affected the precious metals market.

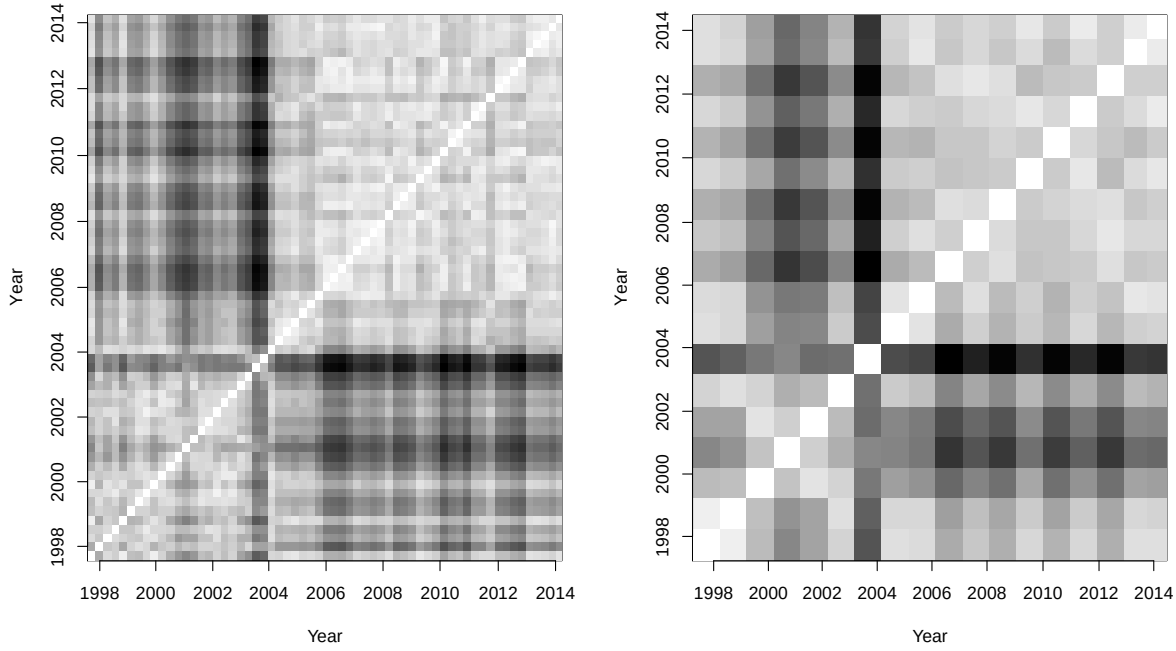


Figure 3 Correlation/similarity matrix for the precious metals market

4 Conclusion

The objective of this study is to analyse the conditional dependence structure among precious metals using the copula-DCC-GARCH methodology and to follow their temporal evolution. The results obtained in the study reveal that the dependence structure is not stable over time. Internal clustering criteria applied to Ward's method prove that two (which seems to be better choice) or three typical patterns of the conditional dependence are plausible. If two market states are assumed, the transition point takes place in April 2004. The state of the precious metals market before and after this moment is stable, with rare and transitory changes. Conditional correlations in the first period are lower than in the second period. If three market states are assumed, the one till April 2004 is stable, but later we observe two patterns which change frequently. The similarity between this two last state are however quite high. Summing up, the results obtained indicate that the dependence structure of precious metals undergoes only one drastic structural change in April 2004. It confirms Sensoy's [22] hypothesis that precious metals will be a single asset class in the near future. A unique opportunity to test this thesis was the global financial crisis, which, however, did not affect the correlation structure of precious metals returns. The results obtained might be of great importance to investors, as they demonstrate that drastic changes of the correlation structure of the precious metals market is currently highly unlikely.

Acknowledgements

Supported by the grant No. 2012/07/B/HS4/00700 of the Polish National Science Centre.

References

[1] Aloui, R., Ben Aïssa, M.S., and Nguyen, D.K.: Conditional dependence structure between oil prices and exchange rates: A copula-GARCH approach. *Journal of International Money and Finance* **32** (2013) 719–738.

[2] Ciner, C.: On the long-run relationship between gold and silver: a note. *Global Finance Journal* **12** (2001), 299–303.

[3] Cochran, S.J., Mansur, I., and Odusami, B.: Volatility persistence in metal returns: a FIGARCH approach. Journal of Economics and Business **64** (2012), 287–305.

[4] Engle, R.: Dynamic conditional correlation. A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journalof Business &EconomicStatistics* **20** (2002), 339–350.

[5] Hammoudeh, S., and Yuan, Y.: Metal volatility in presence of oil and interest rate shocks. *Energy Economics* **30** (2008), 606–620.

[6] Hammoudeh, S.,Yuan,Y., McAleer, M., and Thompson, M.A.: Precious metals-exchange rate volatility

transmissions and hedging strategies. *International Review of Economics & Finance* **19** (2010), 633–647.
[7] Joe H.: *Multivariate models and dependence concepts*, Chapman-Hall, London, 1997.
[8] Lee, T.H., and Long, X.: Copula-Based Multivariate GARCH Model with Uncorrelated Dependent Errors. *Journal of Econometrics* **150** (2009), 207–218.
[9] Li, M., and Yang, L.: Modeling the volatility of futures return in rubber and oil - A Copula-based GARCH model approach. *Economic Modelling* **35** (2013), 576–581.
[10] Lucey, B.M., and Tully, E.: The evolving relationship between gold and silver 1978– 2002: evidence from a dynamic cointegration analysis: a note. *Applied Financial Economics Letters* **2** (2006), 47–53.
[11] Morales, L., Andreosso-O'Callaghan, B.: Comparative analysis on the effects of the Asian and global financial crises on precious metal markets. *Research in International Business and Finance* **25** (2011), 203-227.
[12] Münnix, M. C., Shimada, T., Schäfer, R., Leyvraz, F., Seligman, T. H., Guhr, T., and Stanley, H. E.: Identifying states of a financial market. *Scientific Reports* **2** (2012), Article number: 644, doi:10.1038/srep00644
[13] Nelsen, R.B.: *An Introduction to Copulas*, Springer-Verlag, New York, 1999.
[14] Papież, M., and Śmiech, S.: Causality in mean and variance between returns of crude oil and metal prices, agricultural prices and financial market prices. In: *Proceedings of 30th International Conference Mathematical Methods in Economics. Karviná: Silesian University, School of Business Administration* (Ramík, J. and Stavárek, D., eds.), 2012, 675-680.
[15] Patton, A.J.: Modelling asymmetric exchange rate. *International Economic Review* **47** (2) (2006), 527-556.
[16] Patton, A.J.: A review of copula models for economic time series. *Journal of Multivariate Analysis* **110** (2012), 4-18.
[17] Philippas, D., and Siriopoulos, C.: Putting the 'C' into Crisis: Contagion, Correlations and Copulas on EMU Bond Markets. *Journal of International Financial Markets, Institutions and Money* **27** (2013), 161–176.
[18] Reboredo, J.C.: Is gold a hedge or safe haven against oil price movements? *Resources Policy* **38** (2013), 130–137.
[19] Reboredo, J.C.: Is gold a safe haven or a hedge for the US dollar? Implications for risk management. *Journal of Banking and Finance* **37** (2013), 2665-2676.
[20] Sari, R., Hammoudeh, S., and Ewing, B.T.: Dynamic relationships between oil and metal commodity futures prices. *Geopolitics of Energy* **29 (7)** (2007), 2–13.
[21] Sari, R., Hammoudeh, S., and Soytas, U.: Dynamics of oil price, precious metal prices, and exchange rate. *Energy Economics* **32** (2010), 351–362.
[22] Sensoy, A.: Dynamic relationship between precious metals. *Resources Policy* **38** (2013), 504-511.
[23] Serban, M., Brockwell, A., Lehoczky, J., and Srivastava, S.: Modelling the Dynamic Dependence Structure in Multivariate Financial Time Series, *Journal of Time Series Analysis* **28** (5) (2007) , 763-782.
[24] Silvennoinen, A., and Thorp, S.: Financialization, crisis and commodity correlation dynamics. *Journal of International Financial Markets, Institutions and Money* **24** (2013), 42-65.
[25] Śmiech, S., and Papież, M.: A dynamic analysis of causality between prices on the metals market. In *Proceedings of the International Conference Quantitative Methods In Economics (Multiple Criteria Decision Making XVI)* (Reiff, M. eds.). Bratislava: Slovakia, 2012, 221-225.
[26] Wu, C.C., Chung, H., and Chang, Y.H.: The economic value of co-movement between oil price and exchange rate using copula-based GARCH models. *Energy Economics* **34** (2012), 270–282.
[27] Zolotko, M., and Okhrin, O.: Modelling the general dependence between commodity forward curves. *Energy Economics* **43** (2014),284-296.

The omega ratio in the performance evaluation of mutual funds using the market timing strategy on the Polish financial market

Tomasz Węgrzyn[1]

Abstract. Mutual funds that operate on the Polish financial market are using several investment strategies. Most of them are looking for undervalued stocks. The others are using the market timing strategy in order to profit from changes of the market valuation (as the whole). This kind of mutual funds has to predict future changes in the prices on the financial market. As the result, they are adapting the amount of money invested in stocks to their expectations. The paper concentrates on performance evaluation of mutual funds that are using the market timing strategy.
In order to evaluate the performance of portfolio management several ratios can be used. The most popular is the Sharpe ratio [9]. However, Bertrand and Prigent [4] point out, that it should not be used when investment proportions in asset classes are fluctuating during the time. This situation occurs in case of using the market timing strategy. They suggest using the Omega ratio instead of the Sharpe ratio. In the Omega ratio the whole distribution of returns is used. As the result, in the Omega ratio, the skewness of returns distribution is included while the Sharpe ratio omits it.

Keywords: mutual fund, market timing strategy, omega ratio.

JEL Classification: G11, G14
AMS Classification: 91G10, 91G70

1 Introduction

The results of investments are in the interest of investors. However, they need tools to carry out a performance evaluation. That is why the portfolio performance evaluation is in the interest of professionals (investors) as well as theoreticians. The most popular tool that explains the behavior of the capital market is the Capital Asset Pricing Model introduced by Sharpe [8], Lintner [6] and Mossin [7]. In the CAPM, it is assumed that investors are using the Markowitz model in order to obtain the optimal portfolio of risky assets. Sroczyńska-Baron [10 and 11] shows that it is possible to use the game theory in order to obtain the optimal portfolio. The CAPM shows that when the capital market is in equilibrium then each investor should invest the part of the capital in the market portfolio and the rest of the capital in the risk free asset. As the result, the most effective portfolio is the market portfolio that is represented by the index. As a consequence, the effectiveness of investment is compared with the effectiveness of the market portfolio.

When alternative portfolios are compared then the Sharpe ratio should be used. However, the Sharpe ratio is appropriate when the portfolio structure is constant. It means that the Sharpe ratio is appropriate to portfolio performance evaluation when managers are looking for undervalued stocks to buy them and overvalued stocks to avoid them in the portfolio. Nevertheless, there are some mutual funds that try to implement the market timing strategy. This kind of mutual funds tries to predict future changes in the prices on the financial market. Barczak [1 and 2] shows that Grey Models should be used in order to achieve predictions on the basis of short time series. Such predictions should be used when market timing strategy is implemented. As a result of using the market timing strategy, the part of capital invested in stocks depends on their expectations. When they expect that prices on the capital market will increase then they increase the weight of capital invested in stocks and decrease the weight of capital invested in T-Bills. On the other hand, when they expect that prices on the capital market will decrease then they decrease the weight of capital invested in stocks and increase the weight of capital invested in T-Bills. As a result, the risk in the portfolio is fluctuating during the time. As a consequence, the Sharpe ratio should not be used in the performance evaluation of such mutual funds.

Bertrand and Prigent [4] show that in case of the fluctuation of the asset structure, the Omega ratio should be used. It is because, in the Omega ratio the whole returns distribution is used, while in case of the Sharpe ratio only two moments of distribution are used. The paper concentrates on the performance evaluation of mutual

[1] University of Economics, Department of Applied Mathematics, ul. 1 Maja 50, 40-287 Katowice, tomasz.wegrzyn@ue.katowice.pl

funds that are using the market timing strategy. The results of using the Omega ratio and the Sharpe ratio are compared.

2 The Sharpe ratio and the Omega ratio

The invention of the Capital Asset Pricing Model enabled to find measures that allow to evaluate portfolios' performance. One of such measure is the Sharpe ratio that was introduced in [9]. The Sharpe ratio divides the average portfolio excess return by the standard deviation of returns. As a result it measures the reward (return) to (total) volatility trade-off as it is shown in [5]. As a result, the better portfolio is the portfolio with a higher Sharpe ratio. When the CAPM is valid then the best portfolio in the capital market is the market portfolio. The Sharpe ratio is computed as follows:

$$Sh = \frac{r_p - r_f}{\sigma_p} \tag{1}$$

where:

r_p – average return from the portfolio,

r_f – risk free rate,

σ_p – standard deviation of the returns (total volatility).

The Sharpe ratio is one of many risk-adjusted performance measures for a portfolio. In [5] it is shown that the Sharpe ratio is appropriate measure of the portfolio performance if a given portfolio is the overall portfolio and its composition is not changed during the analyzed period. Otherwise, the Sharpe ratio should not be used.

When the portfolio is actively managed then Bertrand and Prigent [4] point out that performance evaluation made by the Sharpe ratio may be not sufficient. They suggest that in such situation the whole returns distribution should be evaluated. They propose to use the Omega ratio that is computed as follows:

$$\Omega = \frac{\int_L^b (1 - F(X))dx}{\int_a^L F(X)dx} \tag{2}$$

where:

$F(\cdot)$ – cumulative distribution function of returns,

L – break-even point chosen by the investor.

As the formula (2) shows, it is demanded to choose the break-even point L in order to compute the Omega ratio. Choosing the break-even point L, the investor splits the whole return distribution into two parts: returns above the break-even point L are considered as gains while returns below the break-even point L are considered as losses. At a given break-even point L, the investor should always prefer the portfolio with the highest value of the Omega ratio [4]. Moreover, when the Omega ratio is over 1 then there is a higher probability to gain the return over the break-even point L than the return under the break-even point L. It means that investor should prefer only portfolios with the Omega ratio over 1. As Bacmann and Scholz point out [3], the main advantage of the Omega ratio is that it involves all the moments of the return distribution, including skewness and kurtosis. That is why the Omega ratio should be used to evaluate the portfolio performance when the composition of the portfolio is changing during the time.

3 Assumptions and data

There are 5 mutual funds that are implementing market timing strategy included in the study. Chosen mutual funds are investing on the Warsaw Stock Exchange. These mutual funds are as follows:

- TFI Allianz Polska S.A., subfund Allianz Aktywnej Alokacji (later called Allianz);
- Aviva Investors Poland TFI S.A., subfund Aviva Investors Aktywnej Alokacji (later called Aviva);
- PKO TFI S.A., subfund PKO Strategicznej Alokacji-fio (later called PKO);
- Skarbiec TFI S.A., subfund Skarbiec – aktywnej alokacji (later called Skarbiec);
- TFI SKOK S.A., subfund SKOK Aktywny Zmiennej Alokacji (later called SKOK).

The analyzed period is between 1st January 2009 and 31st December 2013. The month logarithmic returns are analyzed. As a risk free rate the WIBOR12M rate (Warsaw Interbank Offered Rate for 12 month period) is assumed[2]. As the benchmark portfolio the WIG index is used.

For each fund the average rate of return, standard deviation of returns, the Sharpe ratio and the Omega ratio are computed.

4 Results

In the table 1, there are returns for analyzed mutual funds achieved in years 2009, 2010, 2011, 2012, 2013 and for the whole analyzed period. In the table 1, there are also returns for the WIG index and the WIBOR12M rate. When returns achieved by each fund are compared with the WIG index it can be noticed:

- the Skarbiec fund does not give at any time a higher return than the WIG index;
- both the SKOK fund and the PKO fund give once a higher return than the WIG index, it is in the year 2011 and the return is negative;
- the Allianz fund gives once a higher return than the WIG index, it is in the year 2011 and the return is positive;
- the Aviva fund gives twice a higher return than the WIG index. In the year 2011 the return is negative, but in the year 2013 the return is positive.

Mutual fund	2009	2010	2011	2012	2013	2009 - 2013
Allianz	19.8%	9.7%	**2.8%**	0.8%	-6.7%	27.0%
Aviva	14.3%	3.5%	**-2.7%**	13.8%	**14.8%**	50.3%
PKO	13.4%	11.0%	**-13.9%**	13.9%	-2.2%	20.7%
Skarbiec	32.8%	-6.8%	-25.0%	20.9%	-5.7%	5.8%
Skok	15.9%	6.8%	**-11.9%**	-1.3%	0.2%	7.9%
WIBOR12M	4.7%	4.4%	4.9%	5.0%	3.0%	24.0%
WIG index	46.9%	18.8%	-20.8%	26.2%	8.1%	88.3%

Table 1 Returns for mutual funds, the WIG index and the WIBOR12M rate

When returns achieved by each fund are compared with the risk free rate (WIBOR12M rate) then it can be noticed:

- the WIG index gives four times a higher return than the WIBOR12M rate;
- the Allianz fund, the Skarbiec fund and the SKOK fund give once a higher return than the WIBOR12M rate;
- both the Aviva fund and the PKO fund give three times a higher return than the WIBOR12M rate.

When the cumulated return achieved by each fund (in the period between 2009 and 2013) are analyzed then it can be noticed:

- the WIG index gives a higher cumulated return than the WIBOR12M rate;
- the Aviva fund gives the highest cumulated return between funds, but it is lower than the cumulated return for the WIG index and it is higher than the cumulated return for the WIBOR12M;
- the second fund on account of the cumulated return is the Allianz fund. It gives the cumulated return between the cumulated return for the WIBOR12M rate and the WIG index;
- the other funds give lower cumulated return than it is for the WIG index as well as the WIBOR12M rate.

In the table 2, there are Sharpe ratios for analyzed mutual funds in years 2009, 2010, 2011, 2012, 2013 and for the whole analyzed period. There are also Sharpe ratios for the WIG index in the table 2. When the Sharpe ratio for each fund is compared with the Sharpe ratio for the WIG index it can be noticed:

- the SKOK fund has not at any time a higher the Sharpe ratio than the WIG index. For the whole analyzed period the Sharpe ratio is negative;
- both the PKO fund and the Skarbiec fund have once (in the year 2009) a higher Sharpe ratio than the WIG index. For the whole analyzed period both funds have negative value of the Sharpe ratio.
- the Allianz fund and the Aviva fund have twice a higher the Sharpe ratio than the WIG index. For the whole analyzed period both funds have positive value of the Sharpe ratio.

[2] In the analyzed period, Polish government did not issue T-Bills (only Bonds are issued).

When mutual funds are evaluated by the Sharpe ratio (and the whole analyzed period is taken into account), then the best fund is the Aviva fund. However, the Sharpe ratio for the Aviva fund is lower than it is for the WIG index. It means that the investment in the Aviva fund was less effective than investment in the WIG index.

Mutual fund	**2009**	**2010**	**2011**	**2012**	**2013**	**2009 - 2013**
Allianz	0.18	**0.43**	**-0.09**	-0.29	-0.29	0.01
Aviva	0.15	-0.04	**-0.21**	0.30	**0.36**	0.11
PKO	**0.35**	0.18	-0.54	0.21	-0.15	-0.02
Skarbiec	**0.45**	-0.29	-0.70	0.27	-0.16	-0.06
Skok	0.19	0.07	-0.62	-0.24	-0.08	-0.08
WIG index	0.31	0.22	-0.45	0.34	0.08	0.12

Table 2 Sharpe ratios for mutual funds and the WIG index

In the table 3, there are Omega ratios for analyzed mutual funds and the WIG index. The omega ratios are computed for chosen break-even points L, that is 0%, 1%, 2%, 4%, 4.3%, 5%. Each rate that is considered as a break-even point is the logarithmic year rate. The break-even point L = 4.3% is equal to the average risk free rate in the analyzed period. If the break-even point L = 0% then there are two mutual funds (the Allianz fund and the Aviva fund) that have a higher Omega ratio than it is for the WIG index. In case of break-even point $L \in \langle 1\%;4\% \rangle$ only the Aviva fund has the Omega ratios higher than the WIG index. It means that if an investor accepts a return rate that is not higher than 4%, the Aviva fund is preferred to the WIG index.

In case of break-even point L = 4.3% the WIG index has the highest Omega ratio. It means that the WIG index is the most efficient if the investor expects a return rate equal at least the risk free rate. However, in this case the Omega ratio is above 1 for the Allianz fund as well as the Aviva fund. It means that those funds are efficient. Nevertheless, they do not earn enough to be more efficient than the WIG index. It suggests that the fees they are charging are too high. The analysis of the Omega ratio for the Aviva fund suggests that if the fee for management was lower by 0.3% a year then the Aviva fund would be more efficient than the WIG index for the risk free rate.

In case of break-even point L = 5% only the Aviva fund and the WIG index have the Omega ratio above 1, what means that they are efficient. Between them, the most efficient is the WIG index – it has the Omega ratio higher than the Aviva fund. Other funds have the Omega ratio under 1. It means that for those funds the probability of getting a higher return than 5% is less than the probability of getting a return lower than 5%.

Mutual fund	**L = 0%**	**L = 1%**	**L = 2%**	**L = 3%**	**L = 4%**	**L = 4.3%**	**L = 5%**
Allianz	**1.586**	1.439	1.306	1.187	1.078	1.047	0.980
Aviva	**1.839**	**1.710**	**1.588**	**1.475**	**1.370**	1.339	1.271
PKO	1.320	1.227	1.139	1.058	0.983	0.961	0.912
Skarbiec	1.059	1.007	0.957	0.910	0.866	0.853	0.823
Skok	1.119	1.039	0.965	0.895	0.830	0.811	0.769
WIG index	1.585	1.528	1.473	1.420	1.369	1.354	1.320

Table 3 Omega ratios for mutual funds and the WIG index

5 Conclusions

In the paper, the efficiency of the mutual funds, that are using the market timing strategy, is examined. The Sharpe ratio and the Omega ratio is used. When for the whole investment period the Sharpe ratio is used, then two funds are efficient (the Allianz fund and the Aviva fund). However they are less efficient than the WIG index. It means that only those two funds earn more than the risk free rate, but they do not earn for the risk.

When the Omega ratio is used and the break-even point is equal to risk free rate than two funds are efficient (the Allianz fund and the Aviva fund). Nevertheless they are less efficient than the WIG index. When the break-even point is equal to 5% then one fund is efficient (the Aviva fund). However it is less efficient than the WIG

index. But, when the break-even point is equal to 4% then the Aviva fund is more efficient than the WIG index. The other efficient fund for the break-even point equal to 4% is the Allianz fund. As a result it can be said that the Aviva fund is the most efficient fund between the chosen funds.

The Omega ratios show that the chosen funds do not earn for fees they are charging. The analysis of the Omega ratio for the Aviva fund shows that if the fee for management was lower by 0.3% a year then the Aviva fund would be more efficient than the WIG index for the risk free rate.

References

[1] Barczak, S.: Zastosowania modeli szarych do prognozowania procesów s kształtnych na przykładzie prognozy na rok akademicki 2011/2012 liczby studentów wszystkich typów studiów w Polsce. In*: Zastosowanie metod ilościowych w zarządzaniu ryzykiem w działalności inwestycyjnej* (Tworek, P., and Barczak, A., eds.). Wyd. UE w Katowicach., Katowice, 2013a, 482-496

[2] Barczak, S.: Zastosowanie teorii szarych systemów do przewidywania przyszłych ofert składanych na aukcjach pierwszej ceny poprzez pryzmat modelu szarego GM(1,1). In*: Studia Ekonomiczne. Innowacje w finansach i ubezpieczeniach – metody matematyczne i informatyczne* (Mika, J., and Dziwok, E., eds.). Wyd. UE w Katowicach., Katowice, 2013b, Vol. 146, 7-18

[3] Bacmann, J.-F., and Scholz S.: Alternative performance measures for hedge funds. *AIMA Journal* **June** (2003).

[4] Bertrand, P., and Prigent, J.: Omega performance measure and portfolio insurance. *Journal of Banking & Finance* **35** (2011), 1811–1823.

[5] Bodie, Z., and Kane, A., and Marcus, A.J.: *Investments*, 8ed., McGraw Hill, 2009.

[6] Lintner, J.: The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics And Statistics* **47** (1965), 13–37.

[7] Mossin, J.: Equilibrium in a capital asset market. *Econometrica* **34** (1966), 768–783.

[8] Sharpe, W.: Capital Asset Prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* **19** (1964), 425–442.

[9] Sharpe, W.: Mutual fund performance. *Journal of Business* **39** (1966), 119–139.

[10] Sroczyńska-Baron, A.: The choice of portfolio based on the theory of cooperative games. In: *European Financial Systems 2013. Proceedings of the 10th International Scientific Conference* (Deev, O., Ksjurova, V., and Krajicek J., eds.). Masaryk University, Brno, 2013a, 305-311.

[11] Sroczyńska-Baron, A.: The analysis of the process of taking over companies based on the theory of games. In: *The 7th International Days of Statistics and Economics, Conference Proceedings* (Löster, T., and Pavelka, T., eds.). Libuše Macáková, MELANDRIUM, Prague, 2013b, 1324-1333

Financial shocks and real exchange rate movements in Poland and Slovakia in the global financial crisis

Justyna Wróblewska[1], Marek A. Dąbrowski[2]

Abstract. Though the Visegrad countries acceded the EU jointly in the spring 2004 they adopted different strategies of monetary integration with the euro area. Slovakia adopted the euro relatively quickly whereas the other countries have not included their currencies in the ERM II yet. Thus, at the time of the global financial crisis the Slovak koruna was pegged and the others had their currencies floating.
Against a backdrop of this difference a set of interesting questions arises: What was the nature of shocks that hit Poland and Slovakia during the crisis? What were the sources of real exchange rate movements? Were real exchange rates acting as a shock absorber or rather a shock-propagating mechanism? Does the ERM II insulate against financial shocks? These questions are addressed by examining the real exchange rates, output, real interest rates and prices in Poland and Slovakia. For both countries we build a set of Bayesian structural VAR models, that allows us to identify supply, demand, monetary and financial shocks. We do not find considerable differences in the structure of shocks hitting Poland and Slovakia in spite of a substantial difference in the exchange rate regime adopted.

Keywords: open economy macroeconomics, real exchange rate, monetary integration, Bayesian structural VAR, financial shocks.

JEL Classification: F41, E44, C11
AMS Classification: 91B64, 91G70

1 Introduction

Exchange rate regimes of Poland and Slovakia have evolved along divergent paths: in the late 1990s both countries had intermediate exchange rate regimes (in 1998 it was even the same regime according to Reinhart and Rogoff [14], i.e. de facto crawling band that is narrower than or equal to +/-5%) but then Poland shifted to free floating whereas Slovakia limited fluctuations of the exchange rate and adopted the euro in 2009. Thus, on the one hand, by the time the recent financial crisis has spread on the global economy, these two countries were at opposite poles of exchange rate arrangement spectrum. On the other hand, Poland and Slovakia are relatively similar economies (i.e. at the comparable level of economic and institutional development, similar history[3]). This creates a unique opportunity to examine whether the exchange rate acted as a shock absorber or rather a shock-propagating mechanism and this is the broad objective of our paper. Specifically, we are interested whether the exchange rate in Poland has been driven by financial shocks before and during the global financial crisis (GFC) to a larger extent than the exchange rate in Slovakia. We raise several related questions: What was the nature of shocks that hit Poland and Slovakia before and during the crisis? What were the sources of real exchange rate movements? Were real exchange rates acting as a shock absorber or rather a shock-propagating mechanism? Does the ERM II insulate against financial shocks?

Empirical evidence on the shock-absorbing property of the flexible exchange rate regime is mixed. Using the SVAR approach Stążka-Gawrysiak [17] and Dąbrowski and Wróblewska [12] show that the exchange rate in Poland was driven by real shocks and thus acted as a shock-absorbing mechanism (see also [11]). In turn Borghijs and Kuijs [8] find that the nominal shocks had significant contribution in the variation of exchange rates of five Central European countries and conclude that the exchange rate appears on average to have acted as much or more as an unhelpful propagator of nominal shocks than as a useful absorber of real shocks. In a more recent study Shevchuk [15] confirms these results.

In a related strand of literature on the resilience of emerging market economies (EMEs) to the GFC is examined by cross-country comparisons. Tsangarides [18] finds that EMEs that peg their currencies weathered the GFC not worse than the those that float but peggers appeared to be faring worse in the recovery period 2010-

[1] Cracow University of Economics, Faculty of Management, Department of Econometrics and Operations Research, Rakowicka 27, 31-510 Cracow, Poland, e-mail: eowroble@cyf-kr.edu.pl.
[2] Cracow University of Economics, Faculty of Economics and International relations, Department of Macroeconomics, Rakowicka 27, 31-510 Cracow, Poland, e-mail: marek.dabrowski@uek.krakow.pl.
[3] We share Blanchard's [5] opinion that 'institutions are central to the workings of a market economy.'

2011. Similarly Blanchard et al. [6] find that there is a little direct effect of the exchange rate regime in limiting the decline in the output growth during the crisis. One of policy lessons drawn by Berkmen et al. [3] after careful examination of factors behind crisis resilience of EMEs is that exchange rate flexibility helped in dampening the impact of the crisis. According to Adler and Tovar [1] exchange rate flexibility mitigates the impact of adverse financial shocks on EMEs, particularly those that are highly financially integrated.

The paper is organised as follows. The economic model, empirical methodology and data are briefly discussed in Section 2. Empirical results are presented in Section 3. Last section concludes.

2 Model, methodology and data

As a theoretical framework for the analysis we use the model developed by Clarida and Galí [9]. Its four building blocks are the IS and LM relations, the uncovered interest rate parity (UIP) condition and the price-setting relation. Deviations of relative output, real exchange rate and relative price from their equilibrium levels can occur due to three structural shocks: supply, demand and monetary.

The model is extended in two ways. First, we relax the uncovered interest parity condition by introducing the risk premium. This allows us to augment the analysis with financial shock which intuitively seems to be important in explanation of exchange rate fluctuations. In order to identify four structural shocks the real interest rate differential is explicitly modelled along with the relative output, the real exchange rate and the relative price. Second, we allow on permanent and transitory components in all shocks.

The long-run identifying restrictions can be derived from the solution of the model and are presented in Table 1 (for details see [9] or [11]). The reaction of the relative price to the financial shock depends on the exchange rate regime: under floating rate the relative price increases and under fixed rate it goes down (see e.g. [10]).

Variable \ shock	Supply	Demand	financial	monetary
Relative output	+	0	0	0
Real interest rate differential	–	+	+	0
Real exchange rate	–	+	–	0
Relative price level	–	+	+/–	+

Table 1 Long-run indentifying restrictions

Artis and Ehrmann [2] point out that the results obtained in any structural VAR exercise are only as good as the identification scheme that is employed to 'make sense' of the residuals. That is why we examine two general classes of SVAR models which differ in the type of restrictions imposed. In the first group there are only zero long-run restrictions from Table 1. In order to identify four shocks we need to introduce one additional zero restriction otherwise one cannot distinguish between demand and financial shocks. Thus, it is assumed that the financial shock has no long-run effect on the real interest rate differential. In order to identify structural shocks in the first group of models the methodology originally proposed by Blanchard and Quah [7] is employed. The recursive structure of the total impact matrix is obtained with the use of the Cholesky decomposition with the normalizing assumptions based on the diagonal signs displayed in Table 1.

Since the additional zero restriction on the real interest rate differential is somewhat arbitrary we examine the second group of models in which both zero and sign restrictions given in Table 1 are imposed. We do not impose the restriction on the reaction of the relative price level to financial shock since it depends on the exchange rate regime (see [10]). The algorithm proposed by Binning [4] is used to identify structural shocks.

All data are from the Eurostat database. We use quarterly data spanning from 1998:1 to 2013:4. Real GDP is used as a measure of output. Real interest rate is calculated as a difference between 3-month money market nominal interest rate and actual inflation. Real exchange rate is based on average quarterly nominal exchange rate defined as the price of national currency in terms of the euro, so its rise means an appreciation of the domestic currency. Price level is measured with a harmonised index of consumer prices (the same measure is used for inflation). Relative output and relative price level are constructed as the log-differences between domestic and foreign (euro area) variables. Real interest rate differential is the difference between domestic and foreign rates.

3 Empirical results

To completely define the Bayesian VAR models we impose the Normal-Wishart prior structure on the parameters. The matrix Normal distribution is centred around zero with the shrinking covariance structure making

VARs with lower order more preferable. The inverted Wishart distribution is centered around the diagonal matrix $0.01I_4$.

The analysis starts with Bayesian model comparison. For both countries we have compared sets of models which consist of five different VAR specifications, i.e. VAR with 5 throughout 9 lags, a constant and seasonal dummies. Additionally, we have decided to include 0-1 dummy in models for Slovakia to account for the participation in the ERM II and the euro area. We assume equal prior probability of each specification. The Savage-Dickey density ratio (see e.g. Verdinelli and Wasserman [19]) has been employed to compute posterior probability of each model specification (Bayes factors comparing each model to a model with only non-stochastic variables on the right sight have been calculated).

For both group of models, i.e. with zero restrictions only and with zero and sign restrictions, and for both countries specification with five lags, i.e. VAR(5) have gained almost all the posterior probability. Thus, the results below are based on VAR(5) models only.

The main findings from the first group of models are not presented in detail because of three reasons. First, the results are quite similar to those from the second group of models. To be sure they are not the same: for example, the contribution of financial shocks to the forecast error variance of the real exchange rate is much higher in both countries than in the second group of models (75% vs. 43% for Slovakia and 79% vs. 38% for Poland at 12-quarter horizon). Second, we have imposed an auxiliary zero restriction in the first group of models for identification purposes. It is not justified by the economic theory and may affect the results. Finally, we want to save some space for historical simulations based on the second group of models.

Below we present the results from the VAR(5) with zero and sign restrictions. Table 2 shows that the variability of the relative output has been mainly driven by supply shocks in both economies: more than 90 per cent of forecast error variance (FEV) is accounted for by these shocks at all horizons in Poland and more than 87 per cent in Slovakia. The other shocks are unimportant in both economies, especially if one takes into account uncertainty related to these estimates. The results are striking: the theory implies that monetary shocks are neutralized in an economy that pegs its currency but not in one that floats. Thus, one could expect the relative output in Poland to be under stronger influence of monetary shocks than in Slovakia. This is not the case.

Forecast	**Proportions of forecast error variance, *h* periods ahead, accounted for by**							
horizon *h*	**supply shocks**		**demand shocks**		**financial shocks**		**monetary shocks**	
				Poland				
1	94.91	(5.49)	2.32	(3.77)	1.90	(3.18)	0.87	(1.43)
4	97.04	(2.88)	1.39	(1.98)	1.12	(1.63)	0.45	(0.62)
8	98.45	(1.46)	0.73	(1.00)	0.59	(0.82)	0.24	(0.32)
12	98.97	(0.96)	0.48	(0.66)	0.39	(0.54)	0.16	(0.21)
				Slovakia				
1	87.89	(10.48)	4.91	(6.35)	4.09	(5.94)	3.11	(4.85)
4	93.77	(5.11)	2.53	(3.02)	2.26	(2.73)	1.44	(1.90)
8	96.74	(2.67)	1.33	(1.56)	1.18	(1.38)	0.74	(0.96)
12	97.84	(1.78)	0.88	(1.04)	0.79	(0.91)	0.49	(0.64)

Notes: 1) posterior standard deviations in parentheses; 2) calculations performed in GAUSS 14, based on 200 thousand accepted MCMC draws after 500 thousand rejected.

Table 2 Posterior mean of forecast error variance decomposition of the realative output in the VAR with zero and sign long-run restrictions

Neither do financial shocks introduce a sharp difference: it rather seems that in the very short-run (one quarter) they find their way into an economy in the real exchange rate variability than in the changes in the relative output. Financial shocks are more important for the variability of the real exchange rate in Poland (38.9%) than in Slovakia (33.5%) (see Table 3). Their contribution to FEV of the relative output is lower in Poland (1.9%) than in Slovakia (4.1%). For the sake of clarity, we do not claim – due to relatively high level of uncertainty – that financial shocks are more important source of the relative output variability in Slovakia than in Poland. Our

point is rather that the opposite hypothesis is even less justified and therefore one should not argue that flexible exchange rate regime in Poland acts as a propagation mechanism of financial (or monetary) shocks.

Real shocks (supply and demand) have non-negligible contribution (circa 60%) to the variability of the real exchange rate both in Poland and Slovakia (Table 3). It is interesting to observe that in Poland supply shocks, i.e. those that have a permanent impact on output, are twice as important as in Slovakia (13.5% vs. 6.7% at 12-quarter horizon). Moreover, the real exchange rate in Poland fluctuates mainly due to the gyrations of the nominal exchange rate whereas in Slovakia, especially after the euro adoption, its changes are a resultant of movements in the relative price level. These three observations imply that the adjustment mechanism to real shocks rests on changes in the nominal exchange rate in Poland and on changes in the relative price level in Slovakia. Detailed results on the FEV decomposition lends support to this hypothesis: total contribution of real shocks to the variability of the relative price level (not shown in tables) is 9.4 per cent in Poland and 16 per cent in Slovakia (at 12-quarter horizon).

Since monetary shocks are unimportant to changes in the real exchange rate, its remaining variability is accounted for by financial shocks. Similarly to the results in [13] we find that these shocks are quite important and the identification procedure that does not allow on them is at risk of misrepresenting the contributions of other shocks. Financial shocks are almost equally important for real exchange rate variability in Poland and Slovakia. Therefore, one cannot argue that the fixed exchange rate regime insulates against financial shocks: in such a case they simply work out through the relative price level and not through the nominal exchange rate. Again, their contribution to the variability of the relative price level (not shown in tables) is substantially smaller in Poland than in Slovakia (3.5% vs. 18.5% at 12-quarter horizon).

Forecast	**Proportions of forecast error variance, *h* periods ahead, accounted for by**							
horizon *h*	**supply shocks**		**demand shocks**		**financial shocks**		**monetary shocks**	
				Poland				
1	13.15	(12.66)	46.56	(28.90)	38.90	(28.60)	1.39	(2.21)
4	13.45	(12.94)	47.96	(28.91)	38.00	(28.54)	0.58	(0.88)
8	13.50	(13.18)	48.29	(29.03)	37.91	(28.63)	0.31	(0.46)
12	13.53	(13.33)	48.40	(29.10)	37.87	(28.69)	0.21	(0.31)
				Slovakia				
1	4.76	(6.55)	59.59	(25.00)	33.46	(24.56)	2.19	(3.47)
4	5.65	(6.90)	52.64	(24.86)	40.74	(24.60)	0.97	(1.36)
8	6.37	(7.96)	50.72	(24.98)	42.36	(24.68)	0.55	(0.74)
12	6.65	(8.55)	50.13	(25.16)	42.85	(24.84)	0.37	(0.50)

Notes: see Table 2.

Table 3 Posterior mean of forecast error variance decomposition of the real exchange rate in the VAR with zero and sign long-run restrictions

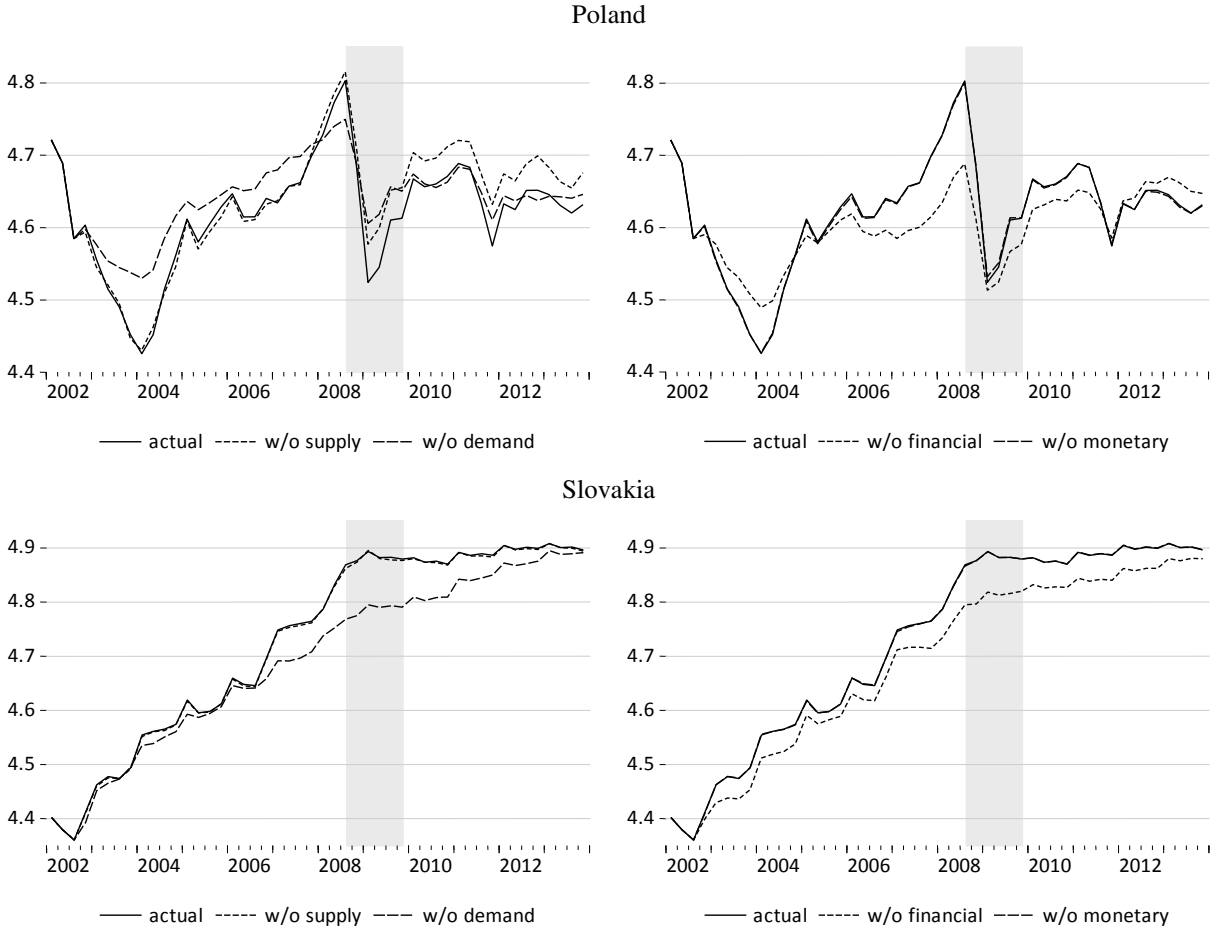
Historical simulations are illustrated in Figure 1. The actual real exchange rate path (solid line) is compared with the path resulting from putting the shock considered equal zero (dashed or dotted line). For instance, in the upper-left panel the observed real exchange rate in Poland is compared with the path without supply shocks (dotted line) and with the path without demand shocks (dashed line). The difference between the solid and dashed/dotted line describes the impact of a given shock on the real exchange rate. Simulations start in 2002:4.

In both economies demand and financial shocks played a crucial role in exchange rate fluctuations. For example, they contributed a lot to the large depreciation of the Polish zloty in the wake of 2001-2002 recession. Monetary shocks were unimportant: paths without these shocks coincide almost completely with the actual paths in both countries. The same is true for supply shocks, although they played some role during and after the global financial crisis in Poland.

Pre-crisis real appreciation was observed in both countries. In Poland it was driven mainly by financial shocks with a positive contribution from demand shocks only at the end of 2007 and in the first half of 2008. In Slovakia contribution of demand shocks to appreciation was not smaller than that of financial shock. It is inter-

esting to observe that in Poland the pre-crisis appreciation was reversed at the beginning of 2009. The real exchange rate of zloty plummeted since financial shocks pushing for the appreciation dissipated and real shocks put a strong downward pressure on zloty. After a rebound in 2009 the exchange rate remained under the influence of supply shocks that were counterbalanced by financial shocks till the debt crisis in the euro area in 2011. Since then both supply and financial shocks have pushed the real exchange rate down.

In contrast to Poland, the pre-crisis appreciation was not reversed in Slovakia. The real exchange rate stabilized since the euro adoption. Unfortunately, its level seems to be overvalued by 7.3 per cent in 2009:1 if the path without demand shocks is treated as a benchmark or by 9.6 per cent for a baseline without financial shocks (other shocks are negligible). Using different methodology Sivák and Péliová [16] argue that the conversion rate of Slovak koruna was set at an overvalued level. By the end of 2013 such a misalignment has shrunk to less than 2 per cent.



Source: own calculations based on 100 000 accepted MCMC draws after 500 000 discarded.

Figure 1 Historical simulations for real exchange rate in Poland and Slovakia

4 Conclusion

The general question that motivates our analysis is whether the exchange rate has acted as a shock absorber or rather a shock-propagating mechanism in Poland and Slovakia. We focus on financial shocks and examine their role in driving the real exchange rate in Poland and Slovakia before and during the global financial crisis.

Even though financial shocks had considerable influence on the real exchange rate in Poland both before and during the global financial crisis, we do not find evidence that exchange rate variability in Poland is dominated by financial (or monetary) shocks to a greater extent than in Slovakia. It is fair to say that in both countries real and financial shocks exert similar influence on the real exchange rate variability. The same is true for the relative output variability which is explained almost only by real shocks in both countries.

Pre-crisis real appreciation has been driven by demand and financial shocks in both economies but in Poland it was reversed at the beginning of 2009 whereas in Slovakia the real exchange rate stabilized at a high level. We also find, consistent with the theory, that real and financial shocks account for a greater part of the relative price

level variability in Slovakia than in Poland. Thus, the adjustment to real and financial shocks works through changes in the relative price level in Slovakia but thorough changes in the nominal exchange rate in Poland.

Overall, we do not find considerable differences in the structure of shocks hitting Poland and Slovakia in spite of a substantial difference in the exchange rate regime adopted. Since forecast error variance decompositions of real exchange rates and relative outputs are similar an interesting extension would be to examine the magnitude of variability in these two variables and a potential trade-off between them (see e.g. [10]). This we leave for further research.

Acknowledgements

The authors gratefully acknowledge financial support from the National Science Centre in Poland (grant no. DEC-2012/07/B/HS4/00723).

References

[1] Adler, G. and Tovar, C. E.: Riding Global Financial Waves: The Economic Impact of Global Financial Shocks on Emerging Market Economies. *IMF Working Paper* **WP/12/188** (2012).

[2] Artis, M., and Ehrmann, M.: The exchange rate – A shock-absorber or source of shocks? A study of four open economies. *Journal of International Money and Finance* **25** (2006), 874-893.

[3] Berkmen, S. P., Gelos, G., Rennhack, R., and Walsh, J. P.: The global financial crisis: Explaining cross-country differences in the output impact. *Journal of International Money and Finance* **31** (2012), 42-59.

[4] Binning, A.: Underidentified SVAR models: A framework for combining short and long-run restrictions with sign-restrictions. *Norges Bank Working Paper* **2013/14** (2013).

[5] Blanchard, O.: Institutions, Markets, and Poland's Economic Performance. *Argumenta Oeconomica Cracoviensia* **6** (2010), 7-11.

[6] Blanchard, O., Faruqee, H., and Das, M.: The initial impact of the crisis on emerging market countries. *Brookings Papers on Economic Activity* (2010), 263–307.

[7] Blanchard, O., and Quah, D.: The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review* **79** (1989), 655-673.

[8] Borghijs, A., and Kuijs, L.: Exchange Rates in Central Europe: A Blessing or a Curse? *IMF Working Paper* **WP/04/2** (2004), 1-28.

[9] Clarida R., and Galí, J.: Sources of Real Exchange Rate Fluctuations: How Important Are Nominal Shocks?, *Carnegie-Rochester Conference Series on Public Policy* **41**, (1994), 1-56.

[10] Dąbrowski, M. A.: Exchange rate regimes and output variability in Central European countries. *Actual Problems of Economics* **2/10** (2012), 80-91.

[11] Dąbrowski, M. A.: Źródła wahań realnych kursów walutowych na Litwie, Łotwie i w Polsce w kontekście kryzysu gospodarczego. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* **269** (2012), 82-96.

[12] Dąbrowski, M. A., and Wróblewska, J.: Przyczyny wahań realnego kursu walutowego w Polsce – wyniki badań z wykorzystaniem bayesowskich strukturalnych modeli VAR. In: Perspektywy integracji gospodarczej i walutowej w Unii Europejskiej w czasach kryzysu. (Opolski, K. and Górski, J., eds.). Wydział Nauk Ekonomicznych Uniwersytetu Warszawskiego, Warszawa, 2013, 119-129

[13] Dąbrowski, M. A., and Wróblewska, J.: Financial shocks as a cause of real exchange rate fluctuations in Poland – evidence from the Bayesian structural VAR models. In: *Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena* (Papież, M., and Śmiech, S., eds.) Foundation of the Cracow University of Economics, Cracow, 2014, 37-46.

[14] Reinhart, C. M., and Rogoff, K. S.: The modern history of exchange rate arrangements: a reinterpretation. *Quarterly Journal of Economics* **119** (2004), 1-48.

[15] Shevchuk, V.: Shock-absorbing properties of the exchange rates in transformation economies: SVAR estimates. In: *Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena* (Papież, M., and Śmiech, S., eds.) Foundation of the Cracow University of Economics, Cracow, 2014, 155-164.

[16] Sivák, R., and Péliová, J.: The Euro and the Slovak Economy in a Period of Financial Crisis. *Argumenta Oeconomica Cracoviensia* **8** (2012), 127-134.

[17] Stążka-Gawrysiak, A.: The Shock-Absorbing Capacity of the Flexible Exchange Rate in Poland. *Focus on European Economic Integration* **4** (2009), 54-70.

[18] Tsangarides, C. G.: Crisis and recovery: role of the exchange rate regime in emerging market countries. *Journal of Macroeconomics* **34** (2012), 470-488.

[19] Verdinelli, I., Wasserman, L.: Computing Bayes factors using a generalization of the Savage-Dickey density ratio, *Journal of the American Statistical Association* **90**, 614-618.

Consumer behaviour of young people from Slovak Republic on the field of the bioproducts by using the correspondence and dependence analysis

Martina Zámková[1], Martin Prokop[2]

Abstract. Organic farming and organic food are terms that attract the attention of not only farmers, but also economists. This paper could address many consumers in the future, thus healthier way of life is becoming more and more popular. This article aims to collect information related to organic farming. It deals with the assessment of shopping behaviour in the organic food market in the Slovak Republic. Respondents from the Slovakia were addressed within the marketing research and potential factors that could affect young people when buying organic products were formulated. There were also established recommendations to retailers of organic products on how to promote the sales of this item to the young generation of Slovaks.
In the questionnaire survey analysis we usually obtain categorical data. Contingency tables are easy way to illustrate the data relations. With respect to the character of the data we use suitable tests of the independence. Correspondence analysis is a multivariate statistical technique. It is conceptually similar to principal component analysis, but applied to categorical rather than continuous data. It provides summarizing a set into two-dimensional graphical form.

Keywords: organic products, marketing research, correspondence analysis, contingency table

JEL Classification: C30
AMS Classification: 62H99

1 Introduction

The main objective of the article is, based on the results obtained from the research, to identify factors that could influence the younger generation to more frequent purchases of organic food. Furthermore, next goal is to formulate recommendations to producers and traders of organic products, on the basis of statistical data processing, how to make young Slovaks become regular customers of bio products for many years.

The article (Vitariusová et al [13]) deals with the research of food intake, leisure time activities and the prevalence of obesity in schoolchildren in Slovakia. Based on the analysis results, the authors recommend increasing consumption of dairy products, fresh fruit and vegetables, whereas BIO quality is a suitable means.

The article of authors (Gomiero, Pimentel and Paoletti [5]) regard this issue from another point of view. Their paper carries out a comparative review of the environmental performances of organic agriculture versus conventional fading. Organic agriculture relies according to the authors on a number of farming practices based on ecological cycles, and aims at minimizing the environmental impact of the food industry, preserving the long term sustainability of soil and reducing to a minimum the use of non renewable resources.

Labelling of organic products

Commission Regulation 271/2010 sets out the graphic appearance of the European Union logo for organic production (see Fig. 1), the conditions for its application and form of the code number of the control institution. The logo is optional for organic products imported into the EU from third countries. The main objective of the European logo is to make organic products easier to be identified by the consumers. Furthermore it gives a visual identity to the organic farming sector and thus contributes to ensure overall coherence and a proper functioning of the internal market in this field. (European commission [4])

Slovak national brand of organic products which certifies organic products in accordance with Act No. 421/2004 Coll. on organic agriculture is shown in Fig. 1. (Ministry of Agriculture SR [10])

[1] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 58601 Jihlava, martina.zamkova@centrum.cz.
[2] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 58601 Jihlava, martin.prokop@vspj.cz.

In the Czech Republic in accordance with Act No. 242/2000 Coll., on organic farming, there results an obligation for manufacturers to label packaged organic food by national logo BIO (biozebra). Any food that is labelled with the word "BIO" or other reference to the way of production in organic farming must also be provided on the packaging by a code of organization that conducted the control whether the product really meets the legal requirements for organic food. Thanks to the code it is possible to trace easily on the pages of the control organizations, whether the product really passed the inspection. (eAGRI [9])

European symbol for product of organic farming	Slovak national product of organic farming	Czech national product of organic farming

Figure 1 Labelling of organic products (eAGRI [9], Ministry of Agriculture SR [10])

2 Materials and Methods

The questionnaire was distributed electronically. Data collection was proceeding from 20st November 2013 to 22st December 2013. More women (68%) than men (32%) participated in the study. Typical respondent's age ranged from 16 to 25 years. Regarding education, respondents were mostly college students and therefore most frequently reported highest level of education was – high school graduation (85%).

It can be stated that this was not a representative sample with respect to the population, but for the fulfilment of the objectives of this article a sample of young respondents is entirely appropriate. It must address especially young people to buy organic food more and more frequently, because young people are potential customers for many years. (Zámková, Prokop [14])

In the questionnaire survey analysis we usually obtain categorical data and easy way to illustrate the data relations are contingency tables. With respect to the character of the data we use suitable tests of the independence. According to (Řezanková [12]) in the case of contingency table of the type $r \times c$ (r is the number of rows, c is the number of columns) we usually use statistics: $\chi^2 = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$. Alternatively $G^2 = \sum_i \sum_j n_{ij} \ln \frac{n_{ij}}{e_{ij}}$, e_{ij} is an expected and n_{ij} real frequency. We use the statistic χ^2 in Pearson´s chi-square test, G^2 in likelihood-ratio test. These two statistics have asymptotically $\chi^2_{(r-1)(c-1)}$ distribution with the presumption of the independence. For further details see (Hindls [8]).

Previous tests can be used in the case of high expected frequencies in the contingency table (more than 5 for each field), see (Hendl [7]). In some studies this rule is not so strict, it is enough to have at most 20% of frequencies less than 5 but all of them more than 1, see (Agresti [1]). According to (Anděl [2]) if frequencies are too small, we can use Fisher´s exact test or we can calculate simulated p-value of χ^2 statistic.

Correspondence analysis (CA) is a multivariate statistical technique. It is conceptually similar to principal component analysis, but applies to categorical rather than continuous data. In a similar manner to principal component analysis, it provides a means of displaying or summarising a set of data in two-dimensional graphical form.

All data should be nonnegative and on the same scale for CA to be applicable, and the method treats rows and columns equivalently. It is traditionally applied to contingency tables - CA decomposes the chi-squared statistic associated with this table into orthogonal factors.

According to (Nenadič, Greenacre [11]), as in principal component analysis, the idea in CA is to reduce the dimensionality of a data matrix and visualize it in a subspace of low-dimensionality, commonly two- or three-dimensional. The data of interest in simple CA are usually a two-way contingency table or any other table of nonnegative ratio-scale data for which relative values are of primary interest. The CA solution was shown by (Greenacre [6]) to be neatly encapsulated in the singular-value decomposition (SVD) of a suitably transformed matrix.

The total variance of the data matrix is measured by the inertia, see, e.g., (Greenacre [6]), which ressembles a chi-square statistic but is calculated on relative observed and expected frequencies.

The weighted sum-of-squares of the coordinates on the *k*-th dimension (i.e., their inertia in the direction of this dimension) is equal to the principal inertia (or eigenvalue), the square of the *k*-th singular value, whereas the standard coordinates have weighted sum-of-squares equal to 1. The implementation of the algorithm follows (Blasius, Greenacre [3]).

The graphical representation of results from CA is commonly done with so-called symmetric maps. In that case, the row and column coordinates on each axis are scaled to have inertias equal to the principal inertia along that axis: these are the principal row and column coordinates. Depending on the situation, other types of display are appropriate. This can be set with the scaling option map in the plotting functions for CA. Software UNISTAT and STATISTICA was used for processing of primary data.

3 Research results

Monthly household income of respondents varies mostly around 800 – 1,600 EUR, while most respondents live in a four-person household. Young people who were addressed do then their purchases mostly in municipalities with a population of 20,000 – 50,000 (26.5%), eventually in towns with 50,000 – 200,000 inhabitants (19.7%). Representation of other groups is relatively balanced, so the recommendations can be directed to all organic food sellers all over the Slovak Republic.

When asked: *What is the organic food in your opinion, what do you recall under this title, what do you associate with this term?* the respondents answered as follows. Respondents most often appropriately marked (almost 87.1%), that they are organically grown products, without chemicals. However, an interesting fact is that more than 41.5% of the respondents associate organic food most often with the fact that the food is expensive. It resulted from the answers to the next question that 79.6% of respondents believe that organic foods are healthier than conventional food and 77.5% think they are of better quality, but only 39% of respondents consider organic food more flavoursome and only 35.4% think they are more attractive than conventional food.

To the question: *Is organic food part of assortment in your favourite store?* 61.9% of respondents answered that yes. The finding that 17% of respondents haven't looked for organic food in the store so far is surprising. Only 45.6% of respondents are persuaded about sufficiently recognizable organic food in stores. Only 35.4% of young people think that organic foods are sufficiently publicized, and surprisingly 11.6% of respondents have not even seen advertising for organic food. If addressed people saw advertising, it was mostly in the magazine and newspaper.

The question: *Do you buy organic food in your household?* respondents answered as follows: it is clear that there is a huge amount of young people who don't buy organic food (40.1%). It is therefore necessary to do basic steps to convince young people that buying organic food makes sense. If respondents don't buy organic food, they stated that the main obstacle for them is the financial aspect – the higher price of organic food (27.9%). Furthermore they stated that they don't believe that organic food is better, non-chemical (15.6%), that it is an unnecessary luxury and no difference (14.3%) and also that organic food is not attractive (12.2%).

If respondents buy organic food, they make their purchases of organic food mostly in hypermarkets and supermarkets (31.3%). Another frequent answer was health food and organic food stores (16.3%), approximately 5.4% of respondents attend farms and farmers markets. A relatively large part of respondents then stated that they grow the organic food by themselves (17.7%). The research also shows that most respondents spend money on organic fruit and vegetables, then for bio milk and dairy products, also for bio flour, cereals and for bio delicacies as well and for meat and sausages. Domestic food production predominates in the available assortment of organic products.

To the question: How often do you buy organic food in your household? respondents most frequently responded that about once per month. And to the question: How much do you approximately spend on buying organic food per year in your household? respondents answered mostly 4 – 20 EUR, respectively less than 4 EUR. It is therefore evident that the Slovaks spend little for organic food per year (if buying), the amount invested in organic food ranges mainly from 0 to 20 EUR.

4 Correspondence and dependence analysis

It is clear from the contingency table of relative frequencies that Slovak women buy organic food slightly more often than men, then it is clear that men quite often don't follow the origin of food, see Tab. 1.

Within the correspondence analysis following questions were assessed: *Do you buy organic food in your household?* and *Have you ever seen an advertising for organic food?* The graph Fig. 2 shows that those respondents, who saw the advertisement for organic food on television, they usually don't buy organic products. Respondents, who saw advertisement in magazines, newspapers or on the Internet, they buy organic food sometimes.

Row relative frequencies	Yes, regularly + Yes sometimes.	No, never.	I don't know, I don't follow whether it is organic food or not.
Woman	63.00%	26.00%	11.00%
Man	53.19%	31.91%	14.89%

Table 1 Contingency table – *Your sex?* and *Do you buy organic food in your household?*

In the graphical output Fig. 3 there are responses to questions: *How often do you buy organic food for your household?* and *How many members are in your household?* The graph shows that the one- to two-person households buy organic food quite often (once a week, several times a month), while larger households with three or more persons buy organic foods less often or not at all.

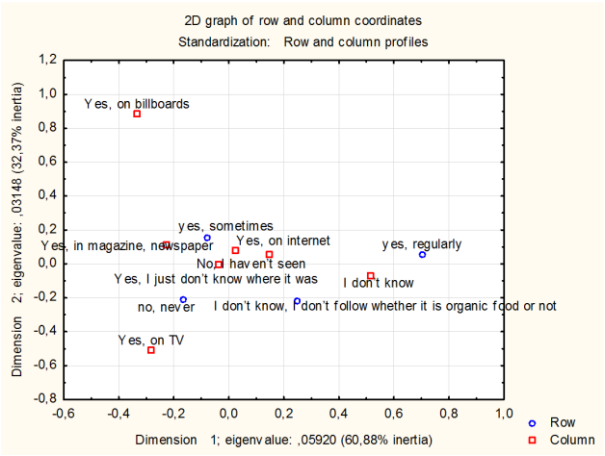


Figure 2 *Do you buy organic food in your household?* and *Have you ever seen an advertising for organic food?*

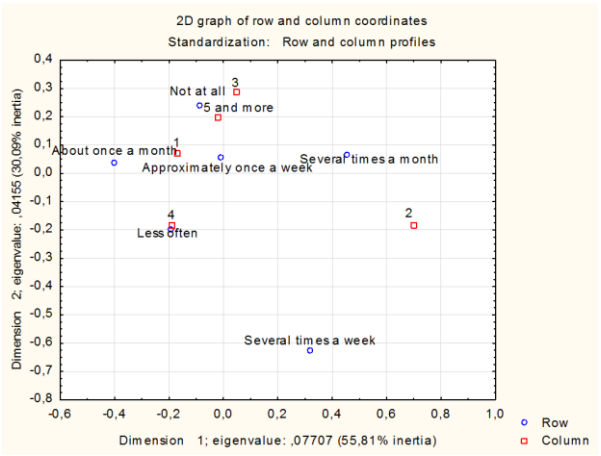


Figure 3 *How often do you buy organic food for your household?* and *How many members are in your household?*

Other responses to questions were monitored: *How much do you spend for organic food for your household per month?* and *Do you think/trust that organic food is - higher quality?* It is evident from correspondence analysis see Fig. 4 that respondents who consider organic food rather better spend maximum of 4 EUR per month. Respondents who spend a larger amount per month (more than 4 EUR) consider organic food definitely better. Respondents, who don't buy organic food, usually don't consider it to be of high quality. When testing the independence the value of the statistics resulted $\chi^2 = 53.04$ with $p < 0.001$. We can talk about the dependence between monthly spending and opinions on quality organic food.

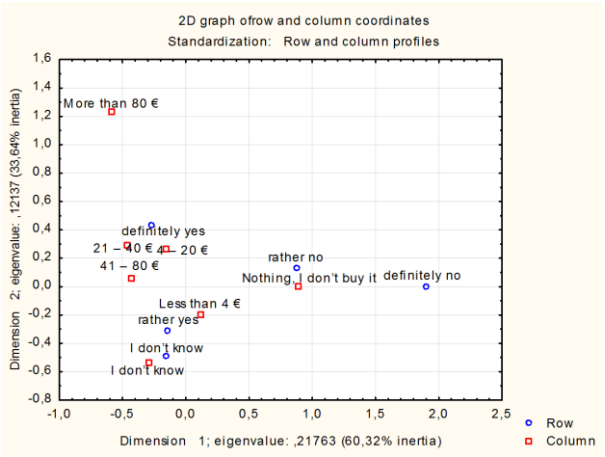


Figure 4 *How much do you spend for organic food for your household per month?* and *Do you think/trust that organic food is - higher quality?*

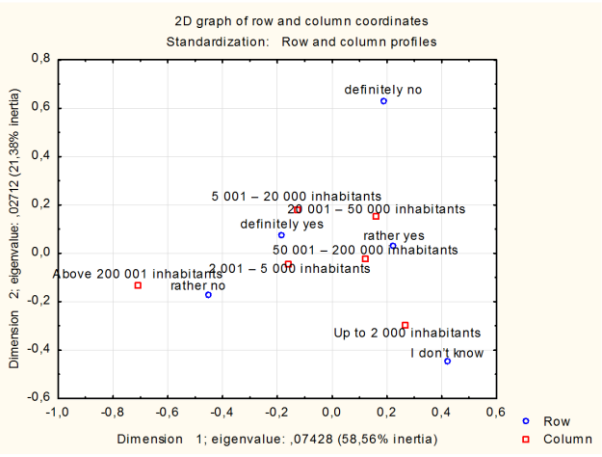


Figure 5 *Do you think/trust that organic food is - higher quality?* and *How many inhabitants are in the municipality where you live and shop?*

Next graphical output resumes answers to questions: *Do you think/trust that organic food is - higher quality?* and *How many inhabitants are in the municipality where you live and shop?* The graph Fig. 5 shows that the residents of the small towns up to 2,000 inhabitants usually don't know whether organic food is better.

Respondents from larger towns (2,000 to 200,000 inhabitants) usually consider organic products as quality. In contrast, residents of large cities over 200,000 inhabitants surprisingly consider organic products as quality rather not. Minimum of respondents answered they certainly don't consider organic products as quality.

Next monitored responses were to questions: *How much do you spend for organic food for your household per month?* and *Do you think/trust that organic food is – healthier?* The correspondence analysis on Fig. 6 shows that those respondents, who trust that organic food is healthier, spend on it about 20 EUR and more per month. And those, who think that organic food is rather healthier, then make the purchase for 20 EUR per month and those, who are not convinced of healthier organic food, don't buy it. When testing the independence the value of the test statistics resulted $\chi^2 = 56.06$ with $p < 0.001$. We can talk about the dependence between monthly spending and opinions on healthier organic food.

Within correspondence analysis there were then assessed answers to questions: *How much do you spend for organic food for your household per month?* and *Do you think/trust that organic food is – tastier?* It is apparent from the graph Fig. 7 that those who strongly consider organic food as tastier spend more than 21 EUR per month on it. Respondents, who are not convinced of tastier organic products, spend less than 4 EUR per month on it. When testing the independence the value of the test statistics resulted $\chi^2 = 37.52$ with $p < 0.05$. We can talk about the dependence between monthly spending and opinions on tastier organic food.

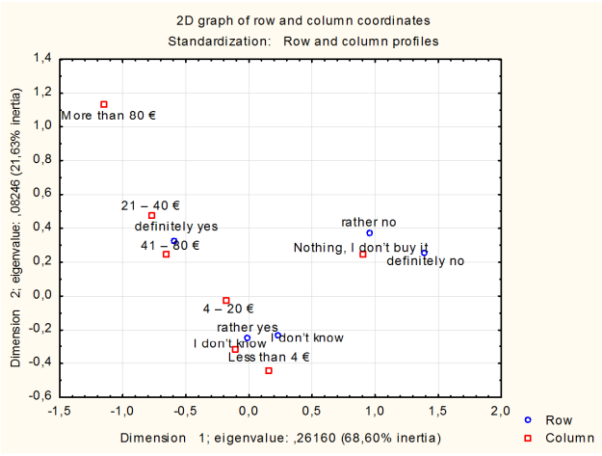


Figure 6 *How much do you spend for organic food for your household per month?* and *Do you think/trust that organic food is – healthier?*

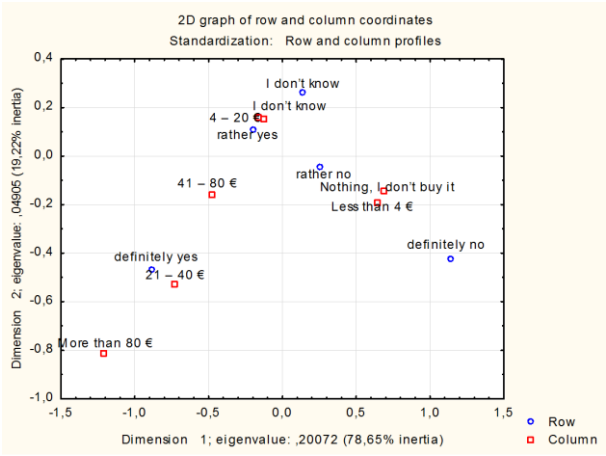


Figure 7 *How much do you spend for organic food for your household per month?* and *Do you think/trust that organic food is – tastier?*

The graph in Fig. 8 describes the responses to the question: What is most important for you while buying a product with this logo? It can be claimed that most of respondents haven't seen the European symbol before. The most common alternatives for the Czech national logo are reactions like: organic products are healthier, organic is better and products are ecologic. Czech national logo was included in the analysis because Slovak students studying in the Czech Republic were addressed too. A large part of respondents couldn't comment Slovak national logo in a similar way.

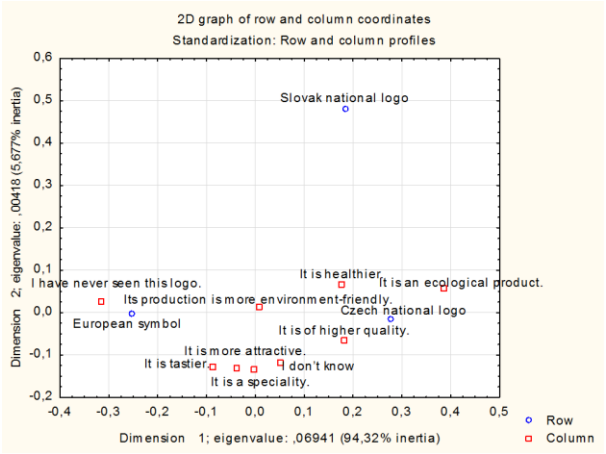


Figure 8 *What is the most important for you when buying a product of organic farming with this logo?*

5 Conclusions

In this paper we tried to assess the shopping behaviour of young respondents from Slovak Republic on field of the organic products. It is clear from the research that there is a considerable amount of young people, who don't buy organic food. In the Slovak Republic there are only 8% of regular buyers of organic products. If respondents don't buy organic food, they stated that the main obstacle for them is the financial aspect – the higher price of organic food. Furthermore they stated that they don't believe that organic food is better, non-chemical. Those, who buy them, regard organic food as healthier than conventional food. It is evident from the research that there is dependence between monthly spending and opinions on healthier, better and tastier organic products.

If respondents buy organic food, they make their purchases of organic food mostly in hypermarkets and supermarkets. The research also shows that most respondents spend money on organic fruit and vegetables, then for bio milk and dairy products. To the question: *How often do you buy organic food in your household?* respondents most frequently responded that about once per month. And to the question: *How much do you approximately spend on buying organic food per year in your household?* respondents answered mostly 4 – 20 EUR, respectively less than 4 EUR.

It is noticeable from the contingency table that Slovak women buy organic food more often than men. From the correspondence analyses it results that it is necessary to focus on more numerous households that buy BIO less frequently or not at all. It is then evident from the correspondence analysis outputs that advertising for organic food on TV is ineffective because of those respondents, who saw the advertisement for BIO on TV, generally responded that they mostly don't buy organic products. They were only addressed by advertising in newspapers, magazines and on the Internet. At the same time it can be stated that the respondents have almost no knowledge of organic food labeling.

It is necessary to take steps enabling us to address the young generation, because these are the customers who will make their purchases for many years. It is then apparent from the analysis that the target group is mainly young women, living in big cities, who don't consider BIO as better, as resulted surprisingly from the research.

The research proved that more than 41% of the surveyed young people associate the term organic food with the food which is expensive. For this reason, we recommend to traders organization of various discount actions also on organic food and to attract young customers on the basis of special discounts on organic food. Similarly, we can recommend organizing more farmers markets and similar events where young people can gain personal experience with organic products and become a fan of them. We also recommend more appropriate advertising campaign highlighting the labeling of organic products and guaranteeing their quality.

References

[1] Agresti, A.: *Categorical Data Analysis*. John Wiley a Sons, New York. 1990.

[2] Anděl, J.: *Základy matematické statistiky*. Matfyzpress, Praha, 2005.

[3] Blasius, J., and Greenacre, M. J.: Computation of Correspondence Analysis. In: *Correspondence Analysis in the Social Sciences* (Greenacre, M. J., Blasius, J., eds.) Academic Press, London, 1994, 53–75.

[4] European commission: *An analysis of EU organic sector* [online]. 2014 [cit. 2014-04-11]. Cited from http://ec.europa.eu/agriculture/organic/eu-policy/data-statistics_cs.

[5] Gomiero, T., Pimentel, D., Paoletti, M. G.: Environmental Impact of Different Agricultural Management Practices: Conventional vs. Organic Agriculture. *Critical reviews in plant sciences* **30** (2011), 95–124.

[6] Greenacre, M. J.: *Theory and Applications of Correspondence Analysis*. Academic Press, London, 1984.

[7] Hendl, J., *Přehled statistických metod: analýza a metaanalýza dat.* Portal, Praha, 2006.

[8] Hindls, R.: *Statistika pro ekonomy*. Professional Publishing, Praha. 2003.

[9] Ministry of Agriculture CR, eAGRI: *Loga pro ekologické zemědělství* [online]. 2013 [cit. 2014-04-15]. Cited from http://eagri.cz/public/web/mze/zemedelstvi/ekologicke-zemedelstvi/loga-a-znaceni/.

[10] Ministry of Agriculture and Rural Development SR: *Ekologické poľnohospodárstvo* [online]. 2014 [cit. 2014-02-04] Cited from http://www.mpsr.sk/index.php?navID=2&navID2=2&sID=43&sID2=45&id=1323.

[11] Nenadič, O., and Greenacre, M., J.: Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. *Journal of Statistical Software*. **20** (2007) 1–13.

[12] Řezanková, H.: *Analýza kategoriálních dat pomocí SPSS.* VŠE, Praha, 2007.

[13] Vitariusová, E. et al: Food intake, leisure time activities and the prevalence of obesity in schoolchildren in Slovakia. *Central European Journal of Public Health* **18** (2010), 192–197.

[14] Zámková, M., and Prokop, M.: Consumers behaviour of students when shopping for organic food in the Czech Republic, *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **61** (2013), 1191–1201.

Mean-risk model optimizing the heavy industrial company's profit with respect to environmental aspects

František Zapletal[1]

Abstract. The paper is devoted to profit maximizing heavy industrial companies' behaviour. This type of companies are limited by many legislative constraints protecting the environment against the negative externalities caused by companies and their large amount of pollution. In this paper, an emphasis is put on the companies in the EU and only on air pollutants. The following legislative tools are involved: emission limits, emission ceilings (caps) and emissions trading system of the EU (EU ETS). A profit optimizing stochastic model respecting mentioned constraints is designed in this paper. The designed model may be used by companies to exploit opportunities arising from legislative factors to get the competitive advantage or at least to minimize a negative influence of the factors on business. Similarly to [5] or [8], prices of emission permit and a demand for products is taken as uncertain. Unlike these works, however, the present paper reacts on new conditions of the EU ETS system and it involves the environmental factors more complexly (e. g. by including emissions limits and caps). The designed model is then verified on the real data of one heavy industrial company.

Keywords: mean-risk model, EU ETS, stochastic programming, ecological constraints.

JEL Classification: C61, Q56
AMS Classification: 90C90

1 Introduction

The aim of the paper is to design an optimization model which will help to choose maximizing profit production structure with respect to both economic and environmental constraints. This model can be used mainly by heavy-industrial company for which a high impact of legislative environmental factors on production is expected and it can be used as a quantitative support for decision-making on production at the tactical level of management. For companies which belong to the carbon leakage sector (the group set by [1] where especially high amount of the carbon dioxide is released, more in Chapter 2.1), the model can be used almost without any changes within the EU. For other companies, slight changes would be required to adapt to different conditions resulting from the emission trading system.

Not too many researches which are devoted to impacts of the emission trading system EU ETS on companies (i. e. from the microeconomic point of view) have been performed. Authors of paper [7] incorporated environmental factors into the company production function without using any method of mathematical programming. Authors of the study [3] designed their deterministic optimization model at the very beginning of the EU ETS system when conditions were quite different and unstable. The study [5] included risk factor in their optimization model but it was designed for energetic companies and for the operational level of management. The closest study to the research of this paper is [8] where multistage stochastic programming model was designed. By its comparing with this research, possibility of using the CER type of allowances (Certified Emission Reduction) is involved and the latest data for verification are used in this paper. On the other hand, the model is designed as only single-stage.

The paper is organised as follows. After this introduction, a general description of environmental factors which are included into the model is performed in the Chapter 2. In the Chapter 3, a list of model assumptions is presented at first. After that, two variants of model are designed and their structure is introduced. The first variant of the model includes only mean values of stochastic variables. The second one is extended by adding the risk measure into the model. The mean-risk optimization model is the result of this extension. The designed models are verified on the real data of one particular steel company. An overview and a description of input data are presented in the Chapter 4. The model verification and its commented results is included in the Chapter 5. The paper is finalized by the Chapter 6 where overall conclusions are presented together with proposals for further research.

[1] VŠB – Technical University of Ostrava, Faculty of Economics, Sokolská 33, Ostrava, CZ, Department of Systems Engineering.

2 Legislative environmental factors influencing the companies' production

Within the EU, many environmental aspects can be identified. All of them follow the same goal – an environmental protection. For purpose of this paper, only the aspects influencing an amount of emissions are taken into account. Factors which regulate an amount of solid and liquid wastes lead to an increase of variable costs of the production. On the other hand emission influencing factors represented mainly by the emission trading, emission caps (ceilings) and limits can directly restrict an amount of a company production.

Moreover, voluntary ecological factors (e. g. EMS – environmental management system which is based on a voluntary reduction of produced wastes, as well as unpredictable factors like production regulation forced by local authorities caused by the exceeded value of air pollution are not involved in this research.

2.1 EU ETS system and its conditions

The main idea of the EU ETS is a fact that each ton of released carbon dioxide by a company must be covered by one emission allowance. For this purpose, EUA allowances (European Union Allowances) were established. A major part of allowances are allocated to companies for free. This applies only for companies within the mentioned carbon leakage sector (e. g. an energy sector does not get any allowance for free) [1]. Additional allowance can be purchased at the secondary market (trading of emission allowances allows selected stock exchanges, e. g. EEX, ICE, KBB, SendeCO2) or at the primary market through the emission auctions. The unused amount of free-allocated allowances can be sold only at mentioned secondary markets. The company has to choose its strategy of the management of emission allowances and it has to decide where and when it purchases/sells allowances during each period of the EU ETS which is set to one year. At the beginning of this period, free allowances are received and in the end of the period all emissions must be covered.

Except of the standard EUA allowance type, other types of allowances can be also used to cover emissions of CO_2. Especially CER allowances (Certified Emissions Reduction) which were established within the Kyoto emission trading system are beneficial for European companies. The CER allowance is many times cheaper than the EUA allowance in the long term. To avoid overusing of CER allowances, limit was set up to 10% of the total company CO_2 emissions [1].

2.2 Emission caps and limits

Emission caps and limits set a maximum allowed amount of selected gasses. Emission cap (ceiling) is a restriction for one year and emission limit restricts an amount of emissions per half an hour. Emission caps are set only for selected sectors or individually for particular companies. Caps and limits are set for CO, NOX, SO2 and airborne dust [2] in the Czech Republic. If the emission cap is set, its value is always smaller than product of emission limit value and a number of half-hours per year (i. e. emission cap is always more restrictive than emission limit when a both are recalculated for the same period).

3 Design of the profit optimizing model

Before a structure of the model is presented, there is a need to summarize its assumptions which are as follows:

- the model is designed as a stochastic programming model with one stage and for only one period whose length is one year;
- the decision is made only at the beginning of the period (no corrections are allowed);
- the production process is homogenous during the entire period (an amount of both production and emissions is the same at any given moment;
- a company generates the profit only by selling its products or by selling unused emission allowances;
- a company always exploits the possibility of using CER allowances to cover 10% of its CO_2 amount;
- only spots of emission allowances are used (although derivatives of the EUA also exist);
- margins for each product are constant in the whole period (selling prices elastically respond to changes in unit variable costs);
- EUA and CER prices are identified as stochastic and mutually independent;
- stochastic independency between demands and allowance prices is supposed;
- values of expected demands are stochastic and mutually dependent;
- exceeding the expected demand is allowed but it is penalized by constant equals to the particular selling price (company must pay cost connected with unsold production but revenue for selling the product is got in one of subsequent periods).

3.1 Stochastic optimization model with the penalization of the objective function without the risk measure

In this version of the model, stochastic variables are substituted only by mean values (no risk measure is involved here). Under this condition together with the assumptions listed at the beginning of the chapter, model of stochastic programming maximizing the total profit of the company with respect to selected environmental factors is defined as follows:

$$\text{maximize} \quad \sum_{i=1}^{n} m_i y_i + \left(R - 0.9 \sum_{i=1}^{n} e_i x_i \right) \cdot E\left(p^{EUA}\right) - 0.1 \sum_{i=1}^{n} e_i x_i \cdot E\left(p^{CER}\right) - \mathcal{P}_i \cdot E\left(y_i - D_i^e\right)^+ - FC \qquad \text{(OF1)}$$

$$\text{s. t.} \quad y_i = \sum_{j=1}^{n} \left(\varepsilon_{ij} - a_{ij}\right) x_j, \quad i = 1,2,...,n \qquad \text{(C1)}$$

$$x_i \le V_i, \quad i = 1,2,...,n \qquad \text{(C2)}$$

$$y_i \ge D_i^c, \quad i = 1,2,...,n \qquad \text{(C3)}$$

$$\sum_{i=1}^{n} h_{ki} x_i \le S_k, \quad k = 1,2,...,m \qquad \text{(C4)}$$

$$\sum_{i=1}^{n} h_{ki} x_i \le 17{,}520 \cdot L_k, \quad k = 1,2,...,m \qquad \text{(C5)}$$

$$x_i \ge 0, \;\; y_i \ge 0, \;\; k = 1,2,...,m \qquad \text{(C6)}$$

Objective function (OF1) describes the total profit of the modelled company (in monetary units) which produce n products each of them in quantity equals to x_i and this quantity is sold in quantity y_i according to demands for *i*-th product. R stands for an amount of free-allocated emission allowances, e_i is an amount of CO_2 emissions released by producing one unit of *i*-th product. Stochastic prices of EUA and CER emission allowances are represented by p^{EUA} and p^{CER}, whereas the last stochastic variable expected amount of demand for *i*-th product is noted as D_i^e. Constant $\mathcal{P}_i$ represents the particular penalization coefficient equals to the selling price for *i*-th product, $(f)^+$ means the positive part of the function f and FC stands for total fixed cost of the company. Constants 0.9 and 0.1 in the equation (OF1) are related to the assumption that company always uses a possibility to cover 10% of its CO_2 emissions by allowances of the CER type. Constraint (C1) ensures a balance between the production and sales and it is derived from the equilibrium of the Leontief input-output model [6]. Coefficient a_{ij} represents a technical coefficient expressing the production connection between the *i*-th and *j*-th product, ε_{ij} is an element of the identity matrix of dimension $n \times n$. Constraints (C2) and (C3) ensure that the amount of production cannot exceed the production capacity V_i for each of n products and sales must cover at least the current demanded amount for the *i*-th product D_i^c (some contracts have been already signed). Next two constraints (C4) and (C5) represent the restrictions given by emission caps S_k and limits L_k for each of m emission types. The coefficient h_{ki} stands for an amount of the *k*-th type of emission released by the production of one unit of the *i*-th product. The constant 17,520 is a number of half-hours in one period (year). The condition (C5) can be expressed in such way because of the one of model assumptions which ensures the homogeneity of the production process. Constraints (C6) provide a non-negativity of model variables.

In designed model, an amount of the production x can be substituted by amount of sales y using the equation (C1). This would lead to the reduction of number of constraints. On the other hand, this substitution would prob-

ably make the model objective function confusing and less transparent. That is why this modification is not further presented.

3.2 Mean-risk optimization model

In this chapter, the model presented in the Chapter 3.1 is extended by adding the risk measures – namely the variance of stochastic variable. All constrains remain the same as in the previous model. Only the objective function is modified by the subtraction of risk resulting from facing the randomness which leads to formulation of the utility function. General expression of the mean-risk model objective function including the risk measure in the form of variance is shown by (1), see e. g. [4].

$$\text{maximize} \qquad U(x) = E(x) - \lambda \cdot VAR(x) \tag{1}$$

Mean-risk version of the designed profit maximizing model is as follows:

$$\text{maximize} \quad \begin{array}{l} \sum_{i=1}^{n} m_i y_i + \mathcal{M} \cdot E\left(p^{EUA}\right) - 0.1 \sum_{i=1}^{n} e_i x_i \cdot E\left(p^{CER}\right) - \sum_{i=1}^{n} \mathcal{P}_i \cdot E(\mathcal{N}) - FC \\ -\lambda \left(\mathcal{M}^2 \cdot Var\left(p^{EUA}\right) + \left(0.1 \sum_{i=1}^{n} e_i x_i\right)^2 \cdot Var\left(p^{CER}\right) + Var\left(\sum_{i=1}^{n} \mathcal{P}_i \cdot \mathcal{N}_i\right) \right) \end{array} \tag{OF2}$$

s. t. (C1), (C2), (C3), (C4), (C5), (C6).

In (OF2), following substitutions were used to keep the reasonable length and transparency: $\mathcal{M} = \left(R - 0.9 \sum_{i=1}^{n} e_i x_i\right)$ and $\mathcal{N}_i = \left(y_i - D_i^e\right)^+$.

The risk resulting from the stochasticity of the p^{EUA}, p^{CER} and D_i^e can be taken separately because of the model assumption that all these three stochastic variables are mutually stochastically independent. The coefficient λ express the decision-maker's attitude to the risk. For the risk-averse decision-maker λ is positive and less than 1. $Var(\cdot)$ stands for the variance and $E(\cdot)$ for the mean value.

The generated model can be still solved using the convex programming – the objective function is still concave in y.

4 Input data

Two main sources of input data were used. The first of them consists of the company internal data (matrix of technical coefficients $a_{ij}; i, j = 1, 2, ..., n$, current amount of demands D_i^c and historical data on demands for products of the company, emission coefficients e_i and h_{ki}, margins m_i, current selling prices $\mathcal{P}_i$, total fixed costs FC and production capacities V_i for $i = 1, ..., n;\ k = 1, ..., m.$

The second part of input data is derived from public a both free and paid databases. Data on historical prices of emission allowances were taken from online available databases of stock exchanges allowing the trading of permits (EEX.com, theICE.com, kbb.sk, sendeCO2). Data on free allocated amount of EUA emission allowances were drawn from paid online available database carbonmarketdata.com. And finally, values for emission caps for the particular type of company were found in [2]. There is no need to involve emission limits because (as it has been already mentioned) caps are always more restrictive.

Researched steel company wants to be kept anonymous. It produces 5 iron and steel products (iron, brams, plates, profiles and cut shapes). Raw iron is used only as a raw material for manufacturing of other products and it is not sold.

With regard to really limited length of the paper, only data on stochastic variables are covered and described in details. Information about probability distributions of all stochastic variables is described in Tab. 1.

As it is mentioned in the Tab. 1, only data from about year period were used to find out the probability distribution of allowance prices. The reason is simple. Because of changes in conditions in the EU ETS in the form of launching of emission auctions, large decline of EUA prices came at the end of the year 2012, see Fig. 1. If the

data from this unstable period are involved, results would be distorted. The end of this period was determined on the base of the breaking regression method performed using the EViews 8 software.

	P^{EUA}	P^{CER}	D(brams)	D(plates)	D(profiles)	D(cut shapes)
Units	EUR	EUR	[thousands of tons]			
Distribution	lognormal	lognomal	truncated multivariate normal			
Mean	6.219	0.5641	20,000	510,000	90,000	28,000
St. dev. (correlation matrix)	0.6461	0.8396	1	0.22	0.66	0.22
			0.22	1	0.22	0.66
			0.66	0.22	1	0.22
			0.22	0.66	0.22	1
Used data frequency	daily	daily	monthly	monthly	monthly	monthly
Used period	02/07/2013 – 01/31/2014		08/2008 – 12/2012			

Table 1 Input data for models and information about probability distributions of stochastic variables

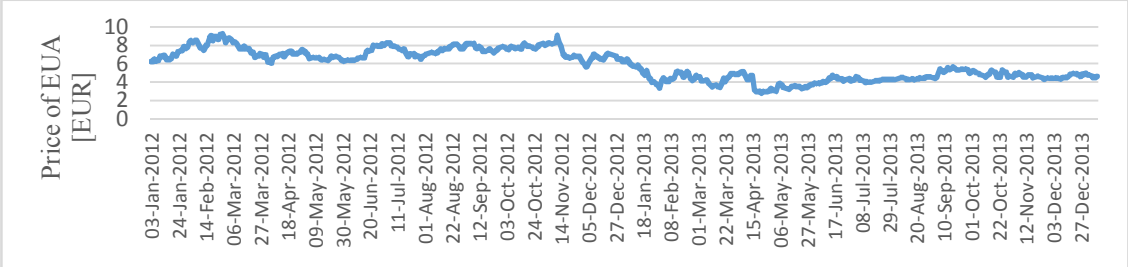


Figure 1 Price of EUA [EUR] on the stock exchange EEX [source: eex.com]

5 Model verification and its result

The designed model was implemented in the MS Excel. While the model without the risk measure was solved analytically, for solving the second model, Monte-Carlo simulation was used to avoid the problems with analytical expressing of variance of the multivariate truncated distribution. After that, the simulated value $Var\sum_{i=1}^{n}\mathcal{P}_i\cdot\left(y_i-D_i^e\right)^+$ were added into the optimization model.

5.1 Results of the stochastic optimization model without the risk measure

Results of the designed stochastic optimization model without using the risk measure are shown in Tab. 2. It can be seen that the company generates the negative profit (i. e. loss). This is caused by current structural problems in the iron and steel industry and persisting effects of the global economic crisis. The value of penalization for exceeding the expected demand was proven to be negligible. A total effect of the EU ETS system is stated by the total profit/loss of allowances and its value is -293.203 EUR. It means that allowances bring additional costs for modelled company (due to the shortage of allowances). This value cannot be neglected by the company management, on the other hand in comparison with the total company profit this additional cost is only minor. Regarding the effects of emission caps and limits, these factors do not influence the production at such low expected demand.

Total margin	Total profit	Penalization	Profit/loss of allowances		Iron	Brams	Plates	Profiles	Cut shapes
[thousands of EUR]				[tons]					
39218.2	-4.063	15.93	-293.203	Production	357251	477650	552760	89994	28000
				Sales	0	19996	509928	89994	28000

Table 2 Results of the stochastic optimization model without involving the risk measure

5.2 Results of the mean-risk model

In view of the results of the model without risk measures (see Tab. 2), the utility function (OF2) was slightly adjusted by excluding the fixed costs FC in order to avoid negative values of the utility function for all lambda coefficients (i. e. for all risk attitudes). This modification induces only the increase of the optimal value by the value of FC. Fig. 2 and Fig. 3 show how the sales volume, total risk and total utility change with changing the risk aversion (lambda). It can be seen that with increasing risk-aversion, final production (and sales) also decrease. Mainly due to fear of exceeding the demand and higher costs caused by emissions trading. According to the Fig. 2, the production structure changes with increasing lambda. Amounts of production of particular commodity converge to values of current demand D_i^c. The convex shape of utility function in lambda is caused especially by effect of production lower bounds D_i^c. If this bounds do not exist, a part of risk connected with allowance price would become growing with sufficiently large lambda coefficient because the decreasing amount of allowances to buy would be replaced by the increasing amount of allowances to sell.

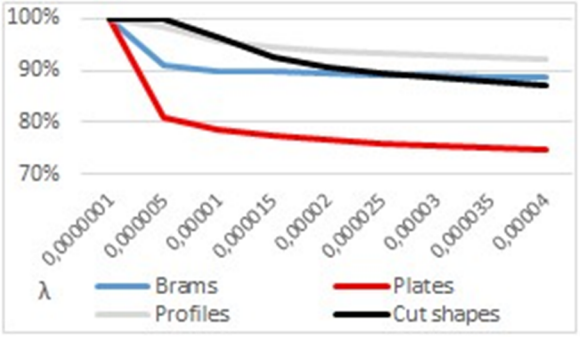


Figure 2 Percentage of sales volume compared to the risk-free optimum

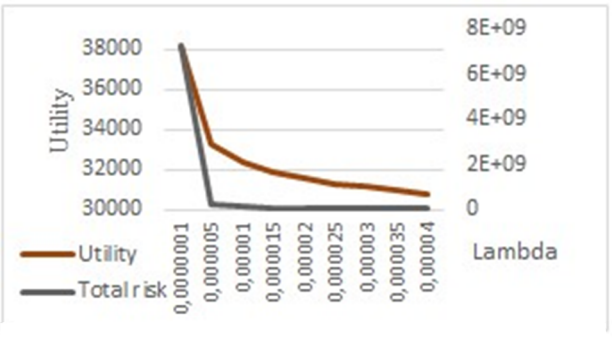


Figure 3 Total risk cause by stochasticity of allowance price and expected demand and optimal utility function values with changing risk aversion

6 Conclusions

The performed research showed the current influence of environmental factors on heavy-industrial companies. Emission caps, limits and especially emission trading were taken into account. Stochastic optimization model and its verification showed that the existence of these environmental factors cannot be ignored. On the other hand, currently, these effects play only a minor role for companies' tactical decision-making which is caused mainly by the persisting depression in the examined steel sector or by too low price of allowances or possibly by still large volume of free-allocated allowances to companies of this sector. That is why a further sensitivity analysis in steel sector and also in other heavy-industrial fields would be very beneficial as well as extending the model to the multistage model. This would help to know the dynamics of the modelled system.

Acknowledgements

This paper was created with financial support of SGS project No.SP2014/146 and the Operational Programme Education for Competitiveness – Project CZ.1.07/2.3.00/20.0296. The support is greatly acknowledged.

References

[1] Rong, A., Landhelma, R.: CO2 emissions trading planning in combined heat and power production via multi-period stochastic optimization. *European Journal of Operational Research* **176** (2007), 1874-1895.
[2] Zhang, B., Xu, L.: Multi-item production planning with carbon cap and trade mechanism. *International Journal of Production Economics* **144** (2013), 118-127.
[3] Directive 2009/29/EC of the European Parliament and of the Council of 23 April 2009 amending Directive 2003/87/EC so as to improve and extend the greenhouse gas emission allowance trading scheme of the Community.
[4] Tang, H., Song, G.: Optimization of Enterprise Production Based on Carbon Emission Credits Trading. *Proceedings of the 2nd International Conference on Green Communications and Networks* (2012).
[5] Lemathe, P., Balakrishnan, N.: Environmental considerations on optimal product mix. *European Journal of Operational Research* **167** (2005), 398-412.
[6] Law No.415/2012 on permitted level of pollution and its detection (Law of the CZ legislative system)
[7] Taha, A. H.: Operations Research: An Introduction. Pearson, 2010. 832 p.
[8] Mansini, R., Ogryczak, W., Speranza, M. G: Twenty years of linear programming based portfolio optimization. *Europan Journal of Operations Research* **234** (2014), 518-535.

Machine Time Scheduling Problems with (max,plus)-Linear Functions

Karel Zimmermann [1]

Abstract. Properties of certain deterministic machine time scheduling problems will be investigated and methods for their solutions will be proposed.The problems will be formulated as optimization problems on image sets of extremally separable functions with additional restrictions on variables. An $O(mn)$-method for solving the problems will be proposed. Modifications and extensions of the problems will be considered. Perspectives of further research will be shortly discussed.

Keywords: machine time scheduling, (max,plus)-linear equation systems.
AMS classification: 90B35, 90C30

1 Introduction.

Algebraic structures with $(\max,+)$-linear functions of the form $f(x_1,\ldots,x_n) = \max_j(c_j + x_j)$ were introduced and applied at first in publications in the sixties and seventies of the 20th century (e.g. [6], [2]). The structures were applied to problems from the area of discrete event systems, machine-time scheduling, reliability theory, transportation problems and others (see e.g. [2], [6]). The recent state of art can be found in monography [1]. Recently appeared several publications dealing with problems on the image set (range) of $(\max,+)$-linear operators (see e.g. [3], [4]). One of the problems consists in approximating a given right hand side vector of an unsolvable $(\max,+)$-linear equation system by the nearest point in the image set of the left hand side $(max,+)$-linear operator of the system (see [3]).

The purpose of this contribution is to generalize the result of [3] using an appropriate parametric approach. In this generalized problem we want to find a solution of a $(\max,+)$-linear equation system with right hand sides $b \in B = \{b \in R^m \ ; \ \underline{b} \leq b \leq \overline{b}\}$, where $\underline{b} \leq \overline{b}$ are given bounds. If $\underline{b} = \overline{b}$, the problem coincides with the problem of [3]. Further it is shown how some precedence relations posed on variables can be easily incorporated in the systems considered. Some motivating examples show a possible application of the results to machine-time scheduling problems.

The following example, the modifications of which can be found in [1], [2] shows a motivation for the research presented in this contribution.

Example 1.
Suppose that n machines $1,2,\ldots,n$ are working in cycles. Let in cycle 1 the machines begin their work at times $x_1,x_2,\ldots,x_n$. We assume that the processing time of machine j is equal to $p_j > 0$. Machine $i,\ 1 \leq i \leq n$ can start the next cycle 2 only after machines $j \in J_i \subseteq \{1,2,\ldots,n\}$ had finished their work in cycle 1. The earliest release times $b_i,\ 1 \leq i \leq n$, at which machine i can begin to work in cycle 2 is computed as follows:

$$b_i = \max_{j \in J_i}(p_j + x_j),\ i = 1,\ \ldots,n.$$

If we set $a_{ij} = p_j$ if $j \in J_i$ and $a_{ij} = -\infty$ otherwise, and $J = \{1,\ \ldots,\ n\}$, we can rewrite the formulae for b_i as follows:

$$b_i = \max_{j \in J}(a_{ij} + x_j),\ i = 1,\ldots,n.$$

The last system is an example of a $(\max,+)$- linear equation system. Properties of such systems were extensively studied in the literature ([1], [2], [6]).

[1]Faculty of Mathematics and Physic, Malostranske namesti 25, Praha 1, Czech Republic, e-mail: karel.zimmermann@mff.cuni.cz

If we prescribe that $b = \hat{b}$, it may happen that the system has no solution. In such situation, we can search the nearest to $\hat{b}$ right-hand side $\tilde{b}$, for which the system is solvable. Such problems were solved in [3].

We can require a less restrictive condition that $b \in B = \{b\ ;\ \underline{b} \leq b \leq \overline{b}\}$, where $\underline{b},\ \overline{b}$ are some chosen bounds. The problem is then to find a $b \in B$, for which the system is solvable and if such b does not exist, to find the nearest to set B right hand side $\tilde{b}$, for which the system can be solved.

Another problem which may occur especially in machine time scheduling is that we may require certain order relation between the starting times x_j in each cycle. If e.g. machine $j+1$ is allowed to work only after machine j had finished its work, i.e. it must be $x_{j+1} \geq x_j + p_j$ if $j,\ j+1 \in J_i$, it arises a question how to incorporate such requirements in our equation system.

In what follows, we propose methods for solving the problems above.

2 A General Problem Formulation.

Let us introduce the following notations. Symbol $y \in R^n$ is supposed to denote a column vector with components $y_j,\ j = 1, \ldots, n$, superscript T denotes the transposition, i.e. $y^T = (y_1, \ldots, y_n)$, A denotes a $(m,n)-$matrix with real entries $a_{ij},\ i \in I \equiv \{1, 2, \ldots, m\},\ J =\equiv \{1, 2, \ldots, n\}$. Let

$$a_i(x) \equiv \max_{j \in J}(a_{ij} + x_j),\ i \in I,\ a(x) = (a_1(x), ..., a_m(x))^T. \tag{1}$$

The set Im(A) will be defined as follows:

$$Im(A) \equiv \{b \in R^m\ ;\ \exists x \in R^n \texttt{ such that } a(x) = b\}. \tag{2}$$

We will study the optimization problem

$$\|b - c\| \quad \longrightarrow \min \tag{3}$$

subject to

$$b \in Im(A),\ c \in B \equiv \{b \in R^m\ ;\ \underline{b} \leq b \leq \overline{b},\} \tag{4}$$

where $\underline{b}, \overline{b}$ are given bounds and $\|b - c\| = \max_{i \in I} |b_i - c_i|$. Let us note that the optimal value of the objective function of this problem is the distance between sets $Im(A)$ and B.

Let t be a real parameter and let

$$M(t) \equiv \{b \in Im(A),\ \underline{b} - t \leq b \leq \overline{b} + t,\} \tag{5}$$

where we set $\underline{b} - t = (\underline{b}_1 - t,\ \ldots,\ \underline{b}_m - t)$ and similarly $\overline{b} + t = (\overline{b}_1 + t,\ \ldots,\ \overline{b}_m + t)$. First we will solve the following auxilliary optimization problem:

$$t \quad \longrightarrow \quad \texttt{subject} \quad \texttt{to } M(t) \neq \emptyset. \tag{6}$$

To solve this problem, we will use a procedure similar to that in [**?**]. Problem (6) can reformulated in the equivalent form

$$t \quad \longrightarrow \quad \min \texttt{ subject to } x \in \tilde{M}(t) \equiv \{x\ ;\ \underline{b} - t \leq a(x) \leq \overline{b} + t\}, \tag{7}$$

where $a(x)$ is defined as in (1). Note that $M(t) \neq \emptyset$ if and only if $\tilde{M}(t) \neq \emptyset$ and both sets are non-empty if t is sufficiently large.

The inequality $a(x) \leq \overline{b} + t$ implies that $\max_{j \in J}(a_{ij} + x_j) \leq \overline{b}_i + t,\ i \in I$ so that it must hold for each $j \in J$

$$a_{ij} + x_j \leq \overline{b}_i + t, \quad \forall i \in I \tag{8}$$

so that

$$x_j \leq \overline{x}_j(\overline{b} + t) \equiv \min_{k \in I}(\overline{b}_k - a_{kj}) + t = \overline{b}_{k(j)} - a_{k(j)j} + t, \quad \forall i \in I. \tag{9}$$

Let us define sets $T_{ij}(t),\ i \in I,\ j \in J$ as follows:

$$T_{ij}(t) \equiv \{x_j\ ;\ \underline{b}_i - a_{ij} - t \leq x_j \leq \overline{x}_j(\overline{b} + t)\}. \tag{10}$$

Let

$$\tau_{ij} = ((\underline{b}_i - a_{ij}) - (\overline{b}_{k(j)} - a_{k(j)j}))/2 \tag{11}$$

Then

$$T_{ij}(t) \neq \emptyset \iff t \geq \tau_{ij}. \tag{12}$$

Let us remark that it follows from (12) and from the definition of τ_{ij} that

$$T_{ij}(t) = \emptyset \iff \underline{b}_i - a_{ij} - t > \overline{x}_j(\overline{b} + t). \tag{13}$$

The following theorem formulates a necessary and sufficient condition for non-emptiness of set $\tilde{M}(t)$ in dependence of t.

Theorem 1. *$\tilde{M}(t) \neq \emptyset$ if and only if*

$$x \leq \overline{x}(\overline{b} + t) \ \ \& \ \ \forall i \in I\ \exists\ j(i) \in J \quad \texttt{such that}\ T_{ij(i)}(t) \neq \emptyset. \tag{14}$$

Proof:

First let us assume that the condition of the lemma is not satisfied. If $x \not\leq \overline{x}(\overline{b} + t)$, then there exists an index $j \in J$ such that $x_j > \overline{x}_j(\overline{b} + t) = \overline{b}_{k(j)} - a_{k(j)j}$ and therefore $a_{k(j)}(x) \geq a_{k(j)j} + x_j > \overline{b}_{k(j)} + t$ so that $x \notin \tilde{M}(t)$. Let now $x \leq \overline{x}(\overline{b} + t)$ and there exists an index $i_0 \in I$ such that $T_{i_0 j}(t) = \emptyset$ for all $j \in J$. Then it must hold for all $j \in J$ (compare (13)) inequalities $\underline{b}_{i_0} - a_{i_0 j} - t > \overline{x}_j(\overline{b} + t)$.Then for an arbitrary $x \in R^n,\ x \leq \overline{x}(\overline{b} + t)$ we have $a_{i_0 j} + x_j < \underline{b}_{i_0} - t$, which implies $a_{i_0}(x) < \underline{b}_{i_0} - t$ so that $x \notin \tilde{M}(t)$.

Let now $\tilde{M}(t) = \emptyset$ and the conditions (13) are satisfied. Then we have $\overline{x}_{j(i)}(\overline{b} + t) \in T_{ij(i)}(t)$ for all $i \in I$ and therefore $\overline{x}(\overline{b} + t) \in \tilde{M}(t)$, which is a contradiction.

□

It follows immediately from Theorem 1 that $\tilde{M}(t) \neq \emptyset$ and therefore also $M(t) \neq \emptyset$ if and only if $t \geq \min_{j \in J} \tau_{ij}$ for each $i \in I$. Therefore if t^{opt} denotes the optimal value of t in problem (6), then

$$t^{opt} = \max_{i \in I} \min_{j \in J} \tau_{ij}. \tag{15}$$

If $t^{opt} \leq 0$, then $a(\overline{x}(\overline{b} + t^{opt})) \in Im(A) \cap B$. If $t^{opt} > 0$, then $Im(A) \cap B = \emptyset$ and $b^{opt} \equiv a(\overline{x}(\overline{b} + t^{opt}))$ is the optimal solution of problem (3), (4). The distance between sets $Im(A)$ and B is equal to $\max(t^{opt}, 0)$. Note that the complexity of computation of t^{opt} is $O(mn)$.

In machine time scheduling problems mentioned in Example 1 requirements concerning an ordering of operation processing may occur. Such requirements are expressed in our general model as inequalities between some variables, e.g. inequality $x_{j+1} \geq x_j + p_j$ in the problem of Example 1 means that if $\{j, j+1\} \subseteq J_i$, then in each cycle of machine i, machine $j+1$ can begin its work only after machine j had finished its work. We will show how such requirements can be incorporated in the general model above.

If $x_{j+1} \geq x_j + p_j,\ p_j \neq 0$ (we admit in the general formulation unlike to Example 1 also $p_j < 0$), we will introduce an additional slack variable s_j and set $x_{j+1} = \max(s_j, x_j + p_j)$ and substitute this relation in the system. We obtain:

$$\max(a_{ij} + x_j, a_{ij+1} + x_{j+1}) = \max(a_{ij} + x_j, a_{ij+1} + \max(s_j, x_j + p_j)),$$

and further

$$\max(a_{ij} + x_j, a_{ij+1} + \max(s_j, x_j + p_j)) = \max(a_{ij} + x_j, \max(a_{ij+1} + s_j, a_{ij+1} + x_j + p_j)),$$

we have further

$$\max(a_{ij} + x_j, \max(a_{ij+1} + s_j, a_{ij+1} + x_j + p_j)) = \max(\tilde{a}_{ij} + x_j, a_{ij+1} + s_j),$$

where

$$\tilde{a}_{ij} = \max(a_{ij}, a_{ij+1} + p_j).$$

We obtain in this way again a $(max, +)$- linear system but instead od variables x_j, x_{j+1} variables x_j, s_j with other coefficients $\tilde{a}_{ij}, p_j$. Note that the number of variables remains unchanged.

In some scheduling problems may be required that the variables are bounded, i.e. $\underline{x}_r \leq x_r \leq \overline{x}_r$ for some $r \in J$. Such restriction can be easily incorporated into our general scheme, namely, we can add to our equation system a new equation $a_{m+1}(x) = b_{m+1}$ and a new restriction $\underline{b}_{m+1} \leq b_{m+1} \leq \overline{b}_{m+1}$, where $a_{m+1}(x) = \max_{j\in J}(a_{m+1j} + x_j)$ with $a_{m+1r} = 0, a_{m+1j} = -\infty$ for $j \in J \setminus \{r\}$, and we set $\underline{b}_{m+1} = \underline{x}_j,\ \overline{b}_{m+1} = \overline{x}_j$ (note that the $-\infty$-values can be replaced by sufficiently low negative numbers). In this case we obtain that for each bound posed on a variable, one additional equation to the original equation system will appear.

We will illustrate the theoretical results by a small numerical example.

Example 2.
Let $m = n = 3$, $\underline{b} = (1, 1, 1)^T,\ \overline{b} = (1, 2, 3)^T$, $I = J = \{1, 2, 3\}$. Matrix A has the form

$$A = \|a_{ij}\| = \begin{pmatrix} 5 & 6 & 1 \\ 4 & 8 & 3 \\ 1 & 3 & 0 \end{pmatrix}.$$

We will find components of $\overline{x}(\overline{b} + t)$:

$$\overline{x}_1 = (\min_{i\in I}(\overline{b}_k - a_{k1}) + t) = \min(1 - 5, 2 - 4, 3 - 1) + t = -4 + t,$$

$$\overline{x}_2 = (\min_{i\in I}(\overline{b}_k - a_{k2}) + t) = \min(1 - 6, 2 - 8, 3 - 3) + t = -6 + t,$$

$$\overline{x}_3 = (\min_{i\in I}(\overline{b}_k - a_{k3}) + t) = \min(1 - 1, 2 - 3, 3 - 0) + t = -1 + t,$$

Let us recall that according to (11) we have for all $i \in I,\ j \in J$

$$\tau_{ij} = ((\underline{b}_i - a_{ij}) - \min_{k\in I}(\overline{b}_k - a_{kj}))/2$$

Therefore we obtain for instance

$$\tau_{11} = ((\underline{b}_1 - a_{11}) - \min_{k\in I}(\underline{b}_k - a_{k1}))/2 = ((1 - 5) - \min(1 - 5, 2 - 4, 3 - 1))/2 = 0,$$

$$\tau_{21} = ((\underline{b}_2 - a_{21}) - \min_{k\in I}(\underline{b}_k - a_{k1}))/2 = ((1 - 4) - \min(1 - 5, 2 - 4, 3 - 1))/2 = 1/2.$$

If we continue the calculations of τ_{ij} in a similar way, we obtain eventually

$$\|\tau_{ij}\| = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 2 & 2 & 1 \end{pmatrix}.$$

so that

$$t^{opt} = \max_{i\in I} \min_{j\in J} = \max(0, -1/2, 1) = 1,$$

$$\overline{x}(\overline{b} + t^{opt}) = (-3, -5, 0)^T,\ a(\overline{x}(\overline{b} + t^{opt})) = (2, 3, 1)^T$$

The distance between $Im(A)$ and B is defined as

$$\rho(Im(A), B) \equiv \min_{b\in Im(A), c\in B} \|b - c\|.$$

We have in this case

$$\rho(Im(A), B) = \max(0, 1) = 1 = \left\|b^{opt} - c^{opt}\right\| = \left\|a(\overline{x}(\overline{b} + t^{opt})) - c^{opt}\right\| = \left\|(2, 3, 1)^T - (1, 2, 1)^T\right\|,$$

where we chose $c^{opt}(1, 2, 1)^T$.

Acknowledgements:

Under the support of GACR 1402424S, Prvouk 202-04/1202

References

[1] Butkovič, P.: Max-linear Systems: Theory and Algorithms, Monographs in Mathematics, Springer Verlag 2010, 271 p.

[2] Cuninghame-Green, R.A.: Minimax Algebra, Lecture Notes in Economics and Mathematical Systems, 166, Springer Verlag, Berlin, 1979.

[3] Krivulin,N.K.: Methods of Idempotent Algebra in Modelling nd Analysis of Complex Systems, S.-Peterburg 2009,255 p. (in Russian).

[4] Krivulin,N.K.: Solution of Generalized Linear Vector Equations in Idempotent Algebra, Vestnik St. Petersburg University: Mathematics 39(1) (2006), pp. 16-26

[5] Litvinov, G.L.Maslov, V.P. Sergeev, S.N. (eds.): Idempotent and Tropical Mathematics and Problems of Mathematical Physics, vol. I, Independent University Moscow, 2007.

[6] Vorobjov, N.N.: Extremal Algebra of positive Matrices, Datenverarbeitung und Kybernetik, 3, 1967, pp. 39-71 (in Russian).

Model of Evaluation of Socio-Economic Disparities of Sub-Regional Units in the Czech Republic

Miroslav Žižka[1]

Abstract. The aim of this paper is to evaluate the disparity of sub-regional units of different hierarchical levels in the Czech Republic. To evaluate the socioeconomic condition of the sub-region factor analysis was used. Source data include 25 social, demographic and economic indicators. Sub-regional units consist of a central municipality to which there are affiliated surrounding catchment municipalities based on the direction of commuting flows to work and school. A center of sub-region must have at least 1,000 inhabitants and basic amenities. Based on the methodical procedure described in the article, there were defined 411 sub-regions in the Czech Republic. Sub-regional units are more suitable for measuring disparities than independent municipalities that represent too detailed level of assessment. On the contrary, higher administrative units are too large to assess disparities; moreover, they are limited by administrative boundaries. Defined sub-regional units are divided according to their function in the territory into four levels and at the same time the differences according to the average value of a factor score are examined.

Keywords: factor analysis, factor loadings, regionalization, socio-economic disparities, sub-regional unit.

JEL Classification: C38, O18, R15
AMS Classification: 62H25

1 Introduction

This article aims to reveal the factors of socio-economic situation of sub-regions in the Czech Republic. Sub-regional units are appropriate territorial units for the evaluation of disparities. As far as their size is concerned, they are between the level of municipalities and administrative districts of municipalities with extended powers. Finding appropriate indicators and methodologies to quantify disparities is important in terms of addressable targeting of regional policy and effective use of public funds in support of economically weak regions.

As stated by Pražák [5], on the one hand, it is more efficient to allocate investments in regions with higher productivity; however, on the other hand, it can lead to economic backwardness of regions with lower productivity and to deepening disparities. Governments are trying to eliminate these disparities by investing a part of the revenues from more developed regions in the infrastructure of less developed regions. Then, the rate of convergence depends on the marginal product of labour, the value of infrastructure and the level of taxation. The efforts to allocate investments in economically stronger regions was empirically confirmed by Felixová [2] who found that the projects financed from the operational programs are not, by priority, directed to regions with a concentrated state support.

On the basis of the law on the support of regional development, regions define in their strategic programmes territorial units that need to be supported with regard to the balanced development of the whole region. As Felixová [3] found in her research, the lower the level of territorial units, the more distinctive resulting image of problems in the given region is, especially if it is related to the whole region in the comprehensive evaluation. On the other hand, this aspect leads to a high degree of fragmentation and consequently to the difficulty in the design of instruments and regional policy measures. Higher administrative units (regions, districts) are too large [4] for the evaluation of disparities, though. Moreover, they are based on an administrative principle. Therefore, functional sub-regions seem to be appropriate territorial units for monitoring disparities.

2 Literary overview

In the literature, there are a number of approaches to measuring disparities of the regions of different hierarchical levels. Most authors deal with the evaluation of higher administrative regions (usually NUTS 3) using established indicators (e.g. GDP per capita, unemployment rate). As an example of this procedure, there can be mentioned a research of Tvrdoň and Skokan [10] who analyzed the disparity in NUTS 3 regions in the Czech Repub-

[1] Technical University of Liberec/Faculty of Economics, Department of Business Administration and Management, Studentská 2, 461 17 Liberec, miroslav.zizka@tul.cz.

lic, Slovak Republic, Hungary and Poland using GDP per capita indicators in the time series of the years 2000-2006. Simple methods of measuring regional disparities such as weighted coefficient of variation, Gini coefficient, Herfindahl index, or normalized Herfindahl index, Ellison and Glaeser index, geographic concentration index or the Theil index are discussed by [8], [9]. The methods were tested on the NUTS 2 and NUTS 3 levels in selected EU countries. Viturka et al. [12] was dealing with the evaluation of the quality of the entrepreneurial and social environment of smaller territorial units – micro-regions of municipalities with extended powers. Methodically, this research was based on the identification of 16 significant factors which were divided into six categories: business, work, infrastructural, local, price and environmental factors. A similar analysis of the dependence of the quality of the business environment on the size of administrative districts of authorized municipalities with extended powers was published in the work of [11]. Černáková and Hudec [1] investigated the quality of life in European cities using 14 variables that were reduced to 5 major components. Cluster analysis was then applied to these components.

In 2011, Žižka [14] performed an analysis of all municipalities in the Czech Republic using 33 indicators and on the basis of factor analysis defined seven significant factors characterizing the socio-economic level. In terms of factor loadings, these factors were interpreted as follows: civic amenities, unemployment, economic activity, domicile attractiveness, settlement population, age structure of inhabitants and branch structure. Municipalities represent the basic territorial units and the building blocks of all regions. However, for the assessment of socio-economic disparities, they are not very useful. The reason lies in the fact that the Czech Republic is characterized by a very heterogeneous structure of municipalities and a very large number of small municipalities (by population) without basic amenities. Hence, comparing communities of different sizes is quite problematic. In addition, it is necessary to consider the fact that in a real situation the amenities can be located in the neighbouring municipality within walking distance. It is, therefore, necessary to assess any municipality in a broader area when taking into account functional linkage between individual municipalities.

When defining functional sub-regions, the criteria from the Rural Development Programme [7] were used. A sub-region consists of the centre (a municipality with minimum of 1,000 residents, post office, primary school, doctor) and catchment communities. Catchment communities are assigned to the centre according to the most significant direction of commuting flows to work and school following the census of 2011.

3 Research methodology

The research was divided into two parts. In the first part, there were defined sub-regions in the Czech Republic and their classification was made. In the second part of the research, there were revealed significant factors of socio-economic situation of sub-region, and at the same time, the analysis of differences in average factor scores between different hierarchical levels of subregions was made.

In more detail, the research process can be divided into the following steps:

A. Definition of sub-regions – regionalization in the Czech Republic. Given that this part is not the topic of the article, there are only presented important points necessary for understanding the process. Detailed procedure is described in [13]. The input file included all 6,251 municipalities in the Czech Republic. Information about the inward and outward commuting between municipalities was obtained from the 2011 census, while they included directional flows of the power of minimum of 10 people.

A1. **Definition of nodal centres** - communities that were targets for commuting for at least one more municipality**.**

A2. Definition of centres of sub-regions - nodal centres which meet all the following criteria: minimum of 1000 inhabitants, post office, primary school at least, general practitioner`s surgery. Nodal centres that do not meet the above conditions were defined as local centres and were assigned to another centre of sub-region in the next step.

A3. Assigning of catchment municipalities to the centre of a sub-region - the surrounding communities were assigned to the centre of the sub-region including local centres according to the most significant total direction of commuting flows of population to work and school.

A4. **Visual analysis** - created sub-regions were shown in the map using GIS tools in order to detect isolated enclaves. To comply with the principle of territorial integrity of the created sub-regions, adjustments were made and some municipalities were assigned to another centre to form a logical unit.

A5. **Classification of sub-regions –** sub-regions fulfil different functions in the area. In the subsequent analysis, it is not appropriate to compare vastly different sub-regions. Centres of sub-regions were therefore classified into four hierarchical levels:

- centre of sub-regional significance – only meets the conditions set out in section A2,

- centre of micro-regional significance – in addition to all the conditions in section A2 it is headquarters of a major employer (minimum of 100 employees),
- centre of regional significance – meets the conditions of the previous category and also it is the seat of a secondary school,
- very significant regional centre – in addition there is an university and it has a minimum of 90 thousand inhabitants.

With the above procedure, there have been defined 411 functional sub-regions, including 49 in the category of sub-regional significance, 109 of micro-regional significance, 242 of regional significance and 11 of very significant regional centres.

B. Evaluation of sub-regional differences in socio-economic field – was carried out using 25 indicators that were discussed with the Czech Statistical Office (further CZSO) in terms of availability and stability of the methodology of monitoring over a longer period of time. The indicators are explained when individual factors are interpreted (see the chapter 4). The evaluation process consists of the following steps:

B1. Assembly of socio-economic indicators for municipalities – data for 2011 were used, characterizing the population, area, age structure, unemployment, economic activity, housing construction, economy, education and healthcare. All data are available from public databases of the CZSO.

B2. Data aggregation and calculation of indicators for sub-regions – absolute data were aggregated for municipalities constituting sub-region and ratio indicators were calculated for all 411 sub-regions.

B3. Verification of the appropriateness of factor analysis – was assessed in two steps: in the first step, there were determined Pearson and partial correlation coefficients and examined the links between different variables. In the second step, there was calculated Kaiser-Melkin-Olkin metric (KMO). The correlation analysis showed that 55% of pair correlations are statistically significant at the 5% significance level. KMO for the whole correlation matrix takes the value of 0.75 and for the individual indicators it ranges from 0.43 to 0.92. KMO values below 0.50 are considered unsatisfactory which concerned only one indicator of secondary education (STRSK), which was therefore excluded from a further analysis. The recalculated value of KMO of the whole matrix was 0.76 and in the individual variables it was higher than 0.50. In view of the amount of KMO of the whole matrix, the appropriateness of using factor analysis can be evaluated as medium-high.

B4. Carrying-out factor analysis – involves determining the number of factors, the choice of methods of extraction and rotation methods. To determine the number of factors, there was used Kaiser's rule - an eigenvalue of factor itself must be higher than one with regard to the percentage of explained variance in the data (requirement of at least 60%). Principal components method was used for extraction of factors. Rotation of factors was carried out using orthogonal Varimax method and oblique Promax method and the results were compared.

B5. Analysis of factor loadings and interpretation of factors – in the analysis it was proceeded from the highest factor loadings of the indicator. Ideally, each indicator should be connected to only one factor. Factor loadings were considered to be significant if it was higher than ± 0.30. Analyzing the factor loadings was followed by the interpretation and naming of factors. It is an activity largely creative.

B6. **Calculation of the factor score of sub-regions** – for each sub-region its factor score was determined. This is a combined measure of all common factors for each sub-region. Factor scores are used to characterize sub-regions and assess the differences between individual sub-regional categories.

All calculations were performed using the statistical software Statgraphics Centurion XVI and OpenStat 11.9.

4 Results of the analysis

Communality or part of the variability of a variable that is explained by common factors ranged between 33 % for the variable of PRLEK up to 94 % for the variable of MRNEZ. At the same time, it is desirable to make the communality values close to 1. The communality that was lower than 60% was identified only in 4 variables. It can therefore be concluded that the used model captures the variability of variables quite well.

Strictly speaking, according to the Kaiser`s rule 8 factors, that explain about 75% of the total variance, should be extracted using the method of principal components. Table 1 shows, however, that the differences between the eigenvalues of factors with serial numbers no. 5 to 8 are very small. According to the shape of the curve of eigenvalues, there is a discontinuity between the third and the fourth factor. According to the Rummel

rule [6], it would be sufficient to define four factors which, however, would explain only 56 % of the variance. The final decision on a number of factors will be taken only after examination of the factor loadings.

Number	Eigenvalue	% of Variance	Cumulative percentage
1	5.361	22.239	22.239
2	4.085	17.020	39.259
3	2.110	8.792	48.051
4	1.909	7.954	56.005
5	1.280	5.334	61.339
6	1.208	5.032	66.371
7	1.071	4.461	70.832
8	**1.009**	**4.202**	**75.034**
9	0.972	4.052	79.086

Table 1 Eigenvalues extracted

Rotation of factors was carried out by Varimax (Var) and Promax (Pro) methods. Table 2 shows the results of extraction of rotated factors with both methods. This table shows that the results are similar, in case of Promax rotation, however, they are more obvious.

Indikator	F1		F2		F3		F4		F5		F6		F7		F8	
Rotation	**Var**	**Pro**	**Var**	**Pro**	**Var**	**Pro**	**Var**	**Pro**	**Var**	**Pro**	**Var**	**Pro**	**Var**	**Pro**	**Var**	**Pro**
AMBUL	0.14	0.02	-0.06	0.00	0.10	0.00	0.21	0.00	0.04	0.00	0.21	0.02	-0.25	-0.01	0.49	0.69
DBYTY	0.45	0.14	0.06	0.00	-0.07	0.00	0.19	0.01	0.15	0.00	-0.04	0.00	0.60	0.25	-0.13	0.00
ESPOP	0.19	0.00	0.14	0.00	0.05	0.00	0.91	0.91	0.21	0.00	-0.02	0.00	0.12	0.00	0.04	0.00
HMROZ	0.03	0.00	0.26	0.01	0.02	0.00	0.11	0.00	0.86	0.85	0.07	0.00	0.00	0.00	0.01	0.00
IEZOB	0.24	0.01	-0.07	0.00	0.78	0.83	-0.05	0.00	-0.01	0.00	0.10	0.00	0.07	0.00	-0.11	0.00
IPMPR	0.08	0.00	0.08	0.00	-0.03	0.00	0.11	0.00	0.90	1.00	0.03	0.00	0.04	0.00	0.02	0.00
ISTAR	0.06	0.00	-0.07	0.00	0.91	1.00	0.09	0.00	-0.01	0.00	-0.12	0.00	-0.09	0.00	0.15	0.00
MATSK	0.02	0.00	0.38	0.07	-0.21	-0.01	0.05	0.00	-0.06	0.00	0.63	0.51	0.06	0.00	-0.01	0.00
MDNEZ	-0.92	-0.99	0.03	0.00	-0.13	0.00	-0.13	0.00	-0.02	0.00	0.05	0.00	0.04	0.00	-0.07	0.00
MRNEZ	-0.95	-1.00	-0.10	0.00	-0.14	0.00	-0.08	0.00	-0.03	0.00	-0.01	0.00	0.04	0.00	-0.09	0.00
MTLPM	-0.94	-0.99	-0.11	0.00	-0.12	0.00	-0.09	0.00	-0.03	0.00	-0.00	0.00	0.05	0.00	-0.10	0.00
OHUZA	0.01	0.00	0.82	0.97	0.01	0.00	-0.13	0.00	0.15	0.00	0.19	0.00	0.01	0.00	-0.06	0.00
OPRVU	0.58	0.28	-0.35	-0.04	0.33	0.03	-0.17	0.00	-0.03	0.00	0.07	0.00	0.17	0.00	-0.26	-0.01
PFARM	0.03	0.00	-0.09	0.00	0.06	0.00	-0.12	0.00	-0.01	0.00	-0.03	0.00	0.27	0.00	0.75	1.00
POBPL	-0.23	0.00	-0.03	0.00	0.00	0.00	0.11	0.00	-0.03	0.00	-0.07	0.00	0.76	1.00	0.20	0.00
PRLEK	-0.07	0.00	0.04	0.00	0.11	0.00	-0.07	0.00	0.09	0.00	0.51	0.75	-0.20	-0.01	0.12	0.00
PVEKC	0.03	0.00	-0.04	0.00	0.90	0.96	0.18	0.00	-0.02	0.00	-0.12	0.00	-0.08	0.00	0.16	0.00
PZAM	0.84	0.86	0.16	0.00	-0.13	0.00	0.23	0.00	0.05	0.00	0.02	0.00	-0.05	0.00	0.05	0.00
SHUZA	-0.12	0.00	0.82	0.93	-0.15	0.00	-0.02	0.00	0.14	0.00	0.17	0.00	0.01	0.00	-0.11	0.00
SPODN	0.20	0.00	-0.04	0.00	0.15	0.00	0.91	1.00	0.05	0.00	-0.03	0.00	0.11	0.00	-0.03	0.00
SUSLU	0.12	0.00	0.82	1.00	-0.03	0.00	0.17	0.00	0.13	0.00	0.00	0.00	-0.02	0.00	0.15	0.00
SUZEM	-0.18	0.00	-0.80	-0.95	0.02	0.00	-0.06	0.00	0.00	0.00	-0.15	0.00	0.00	0.00	0.15	0.00
VOLBY	0.60	0.33	-0.07	0.00	0.50	0.14	0.06	0.00	0.09	0.00	-0.13	0.00	0.13	0.00	-0.19	0.00
ZAKSK	0.00	0.00	0.18	0.00	-0.11	0.00	0.00	0.00	0.05	0.00	0.81	1.00	0.04	0.00	-0.03	0.00

Table 2 Rotated factor loadings

Table 2 shows that the first factor is saturated inversely proportionally by indicators of registered and (MRNEZ) long term (MDNEZ) unemployment and pressure on jobs (MTLPM), directly proportionally by indicator of the proportion of employees to the number of inhabitants (PZAM), it is also weakly related to the proportion of people working on their own account (OPRVU), the intensity of housing construction (DBYTY) and electoral participation (VOLBY). The factor was called **employment.**

Factor 2 is mainly related to the general (OHUZA) and specific density (SHUZA) of population, the share of economic entities in services (SUSLU) and inversely proportionally to the share of employees in agriculture (SUZEM). Thus, it combines in itself the elements of population density and structure of the economy, it also shows certain elements of the rural environment (lower population density and higher proportion of subjects in agriculture are typical for rural regions). **Non-production activities** can be a good name for this factor (i.e. services and agriculture).

Factor 3 is mainly influenced by age index (ISTAR), the average age of the population (PVEKC) and the index of economic burden of population (IEZOB). It was therefore called **age structure.**

The fourth factor consists of the share of economic entities in the population of productive age (ESPOP) and the share of private entrepreneurs in the productive population (SPODN). It can be described as **economic activity of inhabitants.**

The fifth factor is connected with a natural and migration growth (IPMPR) and a gross divorce rate (HMROZ). The factor was called **population.**

The sixth factor is significantly associated with indicators of civic amenities – the number of nurseries (MATSK), primary schools (ZAKSK) and general practitioners per 1000 inhabitants (PRLEK). It was called **civic amenities.**

The seventh factor is saturated by only a single indicator – the average flat living space (POBPL), and in case of Varimax rotation also with the intensity of residential construction. It is, therefore, related to **housing construction.**

The last factor is influenced by indicators of health – the number of outpatient health facilities (AMBUL) and the number of pharmacies (PFARM) to the number of inhabitants. It can be described as **medical amenities.**

Generally, in the factor analysis, the interpretation of factors with a higher sequence number is more difficult. For the further evaluation last two factors were excluded. The reason is the connection with only one indicator, or more precisely the given phenomenon may be explained by other factors (general practitioners in civic amenities factor). The above mentioned six factors explain 66% of the variability in the data.

The last parts of the analysis dealt with the average values of the factor score that depend on the type of subregional centre (see the methodology in Chapter 3, section A5). The average factor score for all factors for all subregions always equals zero. For the own interpretation of the results, it was necessary to specify the type of factor (maximization, minimization). Employment factor (F1) is of a maximization type, i.e. the higher the value of the factor score, the better the situation of a sub-region. Factors F2, F4, F5 and F6 are also of a maximization type. On the contrary, age structure factor (F3) is of the minimization type, i.e. the lower the value of the factor score, the better the situation of a sub-region.

Centre type	F1	F2	F3	F4	F5	F6
sub-region	-1.22	-2.82	0.33	-0.67	-0.69	-0.79
micro-regional	-0.23	-1.47	0.42	-0.62	-0.56	-0.21
regional	0.17	0.91	-0.27	0.23	0.08	0.16
very significant regional	3.96	7.04	0.36	4.15	6.88	2.05

Table 3 Average factor score according to a region type

Lower parameters of employment are characteristic for the centres of a sub-regional type (or higher unemployment rate), non-production activities, economic activities of the population, demographic indicators and worse civic amenities. The age structure is slightly above average. These are mainly centres of rural sub-regions (1,936 people on average).

Micro-regional centres can be assessed by analogy with the previous type, nevertheless, the fact that the average values of individual factor scores are better must be taken into account, even though, they are still rated negatively to the average. Compared to subregional centres, assessment of employment, non-production activities and amenities is significantly better. In particular, the townships and small town (3,078 inhabitants on average) are concerned.

The evaluation of regional centres varies in a relatively narrow band around the average (within one standard deviation). In all factors, the regional centres are rated favourably. These are usually medium-sized towns (13,685 inhabitants on average).

Very important regional centres significantly differ from all previous types with the exception of the age structure. Only in the factor of age structure, they are negatively evaluated. They have significantly above-average evaluation factors of employment, non-production activities, economic activity of the population, demographic metrics of population and civic amenities. These include regional cities with the exception of Jihlava and Karlovy Vary (on average 252,011 inhabitants, without the capital city of Prague 150,333 inhabitants).

5 Conclusion

Using factor analysis, there were defined six significant factors which can be used to evaluate the socio-economic status of sub-regions. In comparison with the analysis conducted in 2011 [14] at the municipal level, this time there are only five factors. The contents of the employment factor, civic amenities, economic activity, and age structure largely correspond. Factors of non-production activity and population are filled differently. In 2011 at the municipal level, there were defined factors of branch structure and settlements which are now "connected" in the factor of non-production activities. Factor of domicile attractiveness at the sub-regional level was not identified at all. Its indicators are represented in three other factors. However, it should be noted that the analysis in 2011 was based on data from 2009.

In terms of practical application the assessment of sub-regions brings benefit in a complex view at a given territorial unit. A municipality is dealt with in terms of the wider area and links to other towns and municipalities. Each centre performs a specific function in the territory. It is; therefore, appropriate to compare centres of the same hierarchical level. For example, if we consider very important centres, we find that although as a whole they are evaluated positively, with the exception of the age structure, in detail, there appear disparities. Specifically, Ostrava and Ústí nad Labem show a negative assessment in employment factor. Even bigger differences were found in age structure factor where half of the very important regional centres are evaluated negatively, but the same number is assessed positively (Olomouc, České Budějovice, Prague, Ústí nad Labem, Liberec and Ostrava). The proposed evaluation methodology can be used to determine the strengths and weaknesses of each sub-region and to formulate measures to alleviate regional disparities. The evaluation of socio-economic disparities and the discovery of related metrics are important in terms of addressed targeting of regional policy instruments.

Acknowledgements

The article was prepared with the support of the project Technology Agency of the Czech Republic ev. No. TD 010029 "Definition of sub-regions for distinguishing and the solution of social and economic disparities".

References

[1] Černáková, V., and Hudec, O.: Quality of Life: Typology of European Cities based on Cluster Analysis. *E&M Economics and Management* **15** (2012), 34–48.
[2] Felixová, K.: Evaluation of the Absorption Intensity of the Entrepreneurial Support in the Regions Funded Intensively by the Government. *E&M Economics and Management* **15** (2012), 17–26.
[3] Felixová, K.: *Podpora podnikání v regionech se soustředěnou podporou státu*. Jan Evangelista Purkyně, Ústí nad Labem, 2012.
[4] Müller, J.: Venkov očima geografa. In: *Sborník ze semináře VENKOV: jeho proměny a územní plánování* (Blatová, T., ed.). Ústav územního rozvoje, Brno, 2005, 4–11.
[5] Pražák, P.: Models of Emergence and Elimination of Regional Economic Disparities as Optimal Control Problems. *E&M Economics and Management* **15** (2012), 15–25.
[6] Rummel, R.J.: *Applied Factor Analysis.* Northwestern University Press, Evanston, 1988.
[7] *Rural Development Program 2007-2013*. Ministry of Agriculture, Prague, 2010.
[8] Tvrdoň, M.: Regional Disparities in Visegrád Group Countries. In: *Proceedings of the 9th International Conference Liberec Economic Forum 2009* (Kocourek, A., ed.). Technical University of Liberec, 2009, 389–397.
[9] Tvrdoň, M.: Regional Income Disparities in the New Member States: Decomposition and Measurement. In: *Proceedings of the 28th International Conference on MATHEMATICAL METHODS IN ECONOMICS 2010, part II.* (Houda, M., and Friebelová, J., eds.). University of South Bohemia in České Budějovice, 2010, 640–645.
[10] Tvrdoň, M., and Skokan, K.: Regional disparities and the ways of their measurement: The case of the Visegrad four countries. *Technological and Economic Development of Economy* **17** (2011), 501–518.
[11] Viturka, M.: Integration Theory of Sustainable Regional Development – Presentation and Application. *Politická ekonomie* **59** (2011), 794–809.
[12] Viturka, M., Wokoun, R., Krejčová, N., Tonev, P., and Žítek, V.: The Regional Relationship between Quality of Business and Social Environment: Harmony or Disharmony? *E&M Economics and Management* **16** (2013), 22–40.
[13] Žižka, M., et al.: *Hospodářský rozvoj regionů*. Professional Publishing, Prague, 2013.
[14] Žižka, M.: Model for Assessment of the Social Economic Level of Municipalities. In: *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011 – part II.* (Dlouhý, M., and Skočdopolová, V., eds.). Professional Publishing, Prague, 2011, 786–791.

32nd International Conference
Mathematical Methods in Economics
MME 2014

Conference Proceedings

Edited by
Jana Talašová
Jan Stoklasa
Tomáš Talášek

Executive editor prof. RNDr. Zdeněk Dvořák, DrSc. et Ph.D.
Responsible editor Mgr. Jana Kreiselová

Cover layout by Jiří Jurečka

All papers passed a blind review process.
Linguistic revision of the texts was not undertaken.
Papers are sorted alphabetically by the names of their authors.

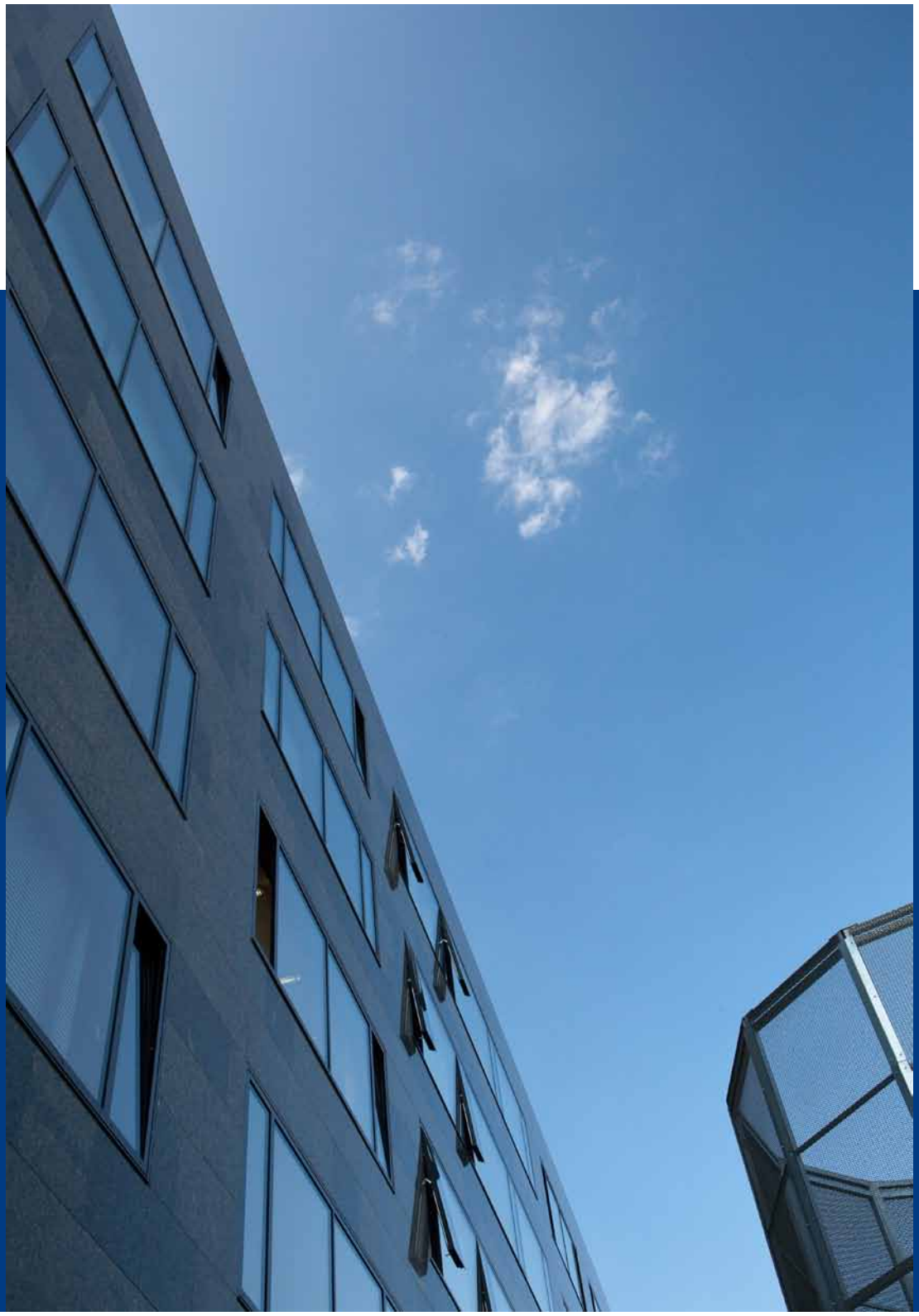
Published and printed by Palacký University, Olomouc
Křížovského 8, 771 47 Olomouc
www.vydavatelstvi.upol.cz
www.e-shop.upol.cz
vup@upol.cz

First edition

Olomouc 2014

VUP 2014/589

ISBN 978-80-244-4209-9



9 788024 442099